

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 22:55:10 ; Search time 42 Seconds
(without alignments)
2439,212 Million cell updates/sec

Title: US-09-804-357B-2

Perfect score: 2129
Sequence: 1 MSVGRRRVKKLLGILMANVF.....RRKPQMIDWISQLQSPNLKC 397

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SP_archaea.*
- 2: SP_bacteria.*
- 3: SP_fungi.*
- 4: SP_human.*
- 5: SP_invertebrate.*
- 6: SP_mammal.*
- 7: SP_mhc.*
- 8: SP_organelle.*
- 9: SP_phase.*
- 10: SP_plant.*
- 11: SP_rodent.*
- 12: SP_virus.*
- 13: SP_vertebrate.*
- 14: SP_unclassified.*
- 15: SP_virus.*
- 16: SP_bacteriap.*
- 17: SP_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2119	99.5	397	11	Q91V18
2	2113	99.2	397	11	Q9Z222
3	1229.5	57.8	406	13	Q8UWM0
4	1176.5	55.3	412	13	Q8UWM3
5	688	32.3	412	13	Q8UWM2
6	658	30.9	397	11	Q8K437
7	655.5	30.8	401	4	Q8NF10
8	653.5	30.6	353	4	Q8N934
9	649	30.5	397	11	Q8K0J2
10	648.5	30.5	388	4	Q8MD21
11	648	30.4	384	4	Q8TDX1
12	648	30.4	418	13	Q8UWM4
13	647.5	30.4	353	4	Q8WY02
14	647.5	30.4	353	4	Q8WNR5
15	647.5	30.4	378	4	Q9C0J1
16	647.5	30.4	378	4	Q8N5W4

17	643.5	30.2	353	4	Q96QH5
18	638.5	30.0	350	11	Q923H4
19	622.5	29.2	372	4	Q9C0J2
20	618	29.0	390	13	Q8UWM1
21	611.5	28.7	377	4	Q96EK0
22	608	28.6	372	11	Q8R0U2
23	605	28.4	372	11	Q9D722
24	602	28.3	370	11	Q8BK98
25	602	28.3	399	11	Q8V116
26	569.5	26.7	277	4	Q8TA24
27	565	26.5	389	11	Q8R319
28	524.5	24.6	374	4	Q8WR86
29	480.5	22.6	326	4	Q9Y526
30	480.5	22.6	326	6	Q9MYM7
31	480.5	22.6	326	11	O54904
32	470	22.1	305	11	Q91V52
33	470	22.1	305	11	Q920V5
34	427.5	20.1	378	4	Q8BYG0
35	408	19.2	409	11	Q920V2
36	405	19.0	409	11	Q91V58
37	405	19.0	422	11	Q8CBX4
38	405	19.0	422	11	Q8BH19
39	404	19.0	409	11	Q91VE9
40	403.5	19.0	376	11	Q8BGY6
41	402	18.9	409	11	Q920V4
42	401	18.8	409	11	Q91V19
43	401	18.8	409	11	Q920V3
44	400.5	18.8	422	11	O54905
45	400	18.8	422	4	O43825

ALIGNMENTS

RESULT 1

Q91V18	PRELIMINARY;	PRT;	397 AA.
AC	Q91V18		
DT	01-DEC-2001 (Tremblrel. 19, Created)		
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)		
DT	01-OCT-2002 (Tremblrel. 22, Last annotation update)		
DE	Beta-1,3-N-acetylglucosaminyltransferase (Beta-1,3-N-acetylglucosaminyltransferase 1).		
DE	B3GNT1.		
GN	Mus musculus (Mouse).		
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C.B17;		
RA	Egan S.E., Cohen B.L., Sarkar M., Ying Y., Cohen S., Singh N.,		
RA	Wang W., Flock G., Goh T., Schachter H.;		
RT	*Molecular cloning and expression analysis of a mouse UDP-		
RT	GlcNAc:Gal(beta1-4)Glc(NAC)-R beta1,3-N-acetylglucosaminyltransferase		
RT	homologous to Drosophila melanogaster Brainiac and the beta1,3-		
RT	galactosyltransferase family*;		
RL	Glycocon]. J. 17:865-872(2000).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C.B17;		
RA	Egan S.E., Cohen B.L., Sarkar M., Ying Y., Cohen S., Singh N.,		
RA	Wang W., Flock G., Goh T., Schachter H.;		
RT	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.		
RL	[3]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Breast tumor;		
RA	Strausberg R.;		
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AY043479; AAK95359.1; -		
DR	EMBL: BC009075; AAH09075.1; -		
DR	MGI: 1889505; B9gnt1.		
DR	InterPro: IPR002659; Glyco_trans_31.		

Q96qh5 homo sapien
Q923h4 mus musculus
Q9c0j2 homo sapien
Q8uwm1 brachydanio
Q96ek0 homo sapien
Q8r0u2 mus musculus
Q9d7z2 mus musculus
Q8bk98 mus musculus
Q8v116 mus musculus
Q8ta24 homo sapien
Q8r319 mus musculus
Q8wr86 homo sapien
Q9y526 homo sapien
Q9ym7 pongo pygma
O54904 mus musculus
Q91v52 mus musculus
Q920v5 mus spicile
Q9by90 homo sapien
Q920v2 mus spicile
Q91v58 mus musculus
Q8cbx4 mus musculus
Q8bh19 mus musculus
Q91ve9 mus musculus
Q8by6 mus musculus
Q920v4 mus musculus
Q91v19 mus musculus
Q920v3 mus musculus
O54905 mus musculus
O43825 homo sapien

DR Pfam: PF01762; Galactosyl_T; 1.
 KW Glycosyltransferase: Transferase.
 SQ SEQUENCE 397 AA: 45883 MW: D8BBAE1866C1D106 CRC64;

Query Match 99.5%; Score 2119; DB 11; Length 397;
 Best Local Similarity 99.5%; Pred. No. 3.1e-186;
 Matches 395; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSVGRRRVLLGLLMMANVFYILVEVSKNSQDKNGKGVIIIPKEKFWKPPSTPRAYWN 60
 DB 1 MSVGRRRVLLGLLMMANVFYILVEVSKNSQDKNGKGVIIIPKEKFWKPPSTPRAYWN 60

QY 61 REQELNRYNPIILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRKDFL 120
 DB 61 REQELNRYNPIILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRKDFL 120

QY 121 LYLRCRNYSLLDQPKCAKPFLLAIAKSLIPHARQAIRRESGRTNNGTQVVRVF 180
 DB 121 LYLRCRNYSLLDQPKCAKPFLLAIAKSLIPHARQAIRRESGRTNNGTQVVRVF 180

QY 181 LLGKTPPEDNHPDLSMDLKFESDKHQDILMWNRYRDTFFNLKVEYLFWRWYSTCPDAEF 240
 DB 181 LLGKTPPEDNHPDLSMDLKFESDKHQDILMWNRYRDTFFNLKVEYLFWRWYSTCPDAEF 240

QY 241 VFKGDDVFNTHHILNLSLSKSKAKDLFGDVYHNGAPHRDKKLYIPEVFTGYV 300
 DB 241 VFKGDDVFNTHHILNLSLSKSKAKDLFGDVYHNGAPHRDKKLYIPEVFTGYV 300

QY 301 PPYAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKLGVLVPEKHKGFTDIEE 360
 DB 301 PPYAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKLGVLVPEKHKGFTDIEE 360

QY 361 KNKKNICSVIDMLVHSRKPQEMIDWSQLSPNLK 397
 DB 361 KNKKNICSVIDMLVHSRKPQEMIDWSQLSPNLK 397

RESULT 2
 Q92222 PRELIMINARY; PRT: 397 AA.

ID Q92222
 AC Q92222;
 DT 01-MAR-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE N-acetylglucosaminide beta-1,3-N-acetylglucosaminyltransferase
 DE (EC 2.4.1.149) (Poly-N-acetylglucosamine extension enzyme) (N-acetylglucosaminyltransferase) (Beta3GNT).
 GN B3GNT1 OR BETA3GNT.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RC STRAIN-ICR; TISSUE-Neonatal brain;
 RX MEDLINE=99110903; PubMed=9892646;
 RA Zhou D., Dinter A., Gutierrez Gallego R., Kamerling J.P.,
 Vliegdenhart J.F.G., Berger E.G., Hennet T.;
 RT "A beta-1,3-N-acetylglucosaminyltransferase with poly-N-
 acetylglucosamine synthase activity is structurally related to beta-
 1,3-galactosyltransferases.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:406-411(1999).
 RN [2]
 RP REVISIONS.
 RA Zhou D., Berger E.G., Hennet T.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: CAN INITIATE THE SYNTHESIS OR THE ELONGATION OF THE
 CC LINEAR POLY-N-ACETYLGLUCOSAMINOGLYCANS.
 CC -!- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-
 CC GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R = UDP + N-ACETYL-BETA-D-
 CC GLUCOSAMINYL-1,3-BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R.
 CC -!- COFACTOR: MANGANESE.
 CC -!- PATHWAY: GLYCOSYLATION.

CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, LUNG, KIDNEY AND
 CC TESTIS AND TO A LESSER EXTENT IN LIVER AND SKELETAL MUSCLE. NO
 CC EXPRESSION IN SPLEEN.
 DR EMBL: AF092050; AAD09763.2; -.
 DR MGD: MGI:1889505; B3gnt1.
 DR InterPro: IPR002659; GlycoTrans_31.
 DR Pfam: PF01762; Galactosyl_T; 1.
 KW Transferase: Glycosyltransferase; Transmembrane; Signal-anchor;
 KW Glycoprotein; Golgi stack.
 FT DOMAIN 1 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 8 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT DOMAIN 29 325 LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 30 30 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 397 AA: 45813 MW: 29BBADF7B70A6B CRC64;

Query Match 99.2%; Score 2113; DB 11; Length 397;
 Best Local Similarity 99.2%; Pred. No. 1.1e-185;
 Matches 394; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSVGRRRVLLGLLMMANVFYILVEVSKNSQDKNGKGVIIIPKEKFWKPPSTPRAYWN 60
 DB 1 MSVGRRRVLLGLLMMANVFYILVEVSKNSQDKNGKGVIIIPKEKFWKPPSTPRAYWN 60

QY 61 REQELNRYNPIILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRKDFL 120
 DB 61 REQELNRYNPIILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRKDFL 120

QY 121 LYLRCRNYSLLDQPKCAKPFLLAIAKSLIPHARQAIRRESGRTNNGTQVVRVF 180
 DB 121 LYLRCRNYSLLDQPKCAKPFLLAIAKSLIPHARQAIRRESGRTNNGTQVVRVF 180

QY 181 LLGKTPPEDNHPDLSMDLKFESDKHQDILMWNRYRDTFFNLKVEYLFWRWYSTCPDAEF 240
 DB 181 LLGKTPPEDNHPDLSMDLKFESDKHQDILMWNRYRDTFFNLKVEYLFWRWYSTCPDAEF 240

QY 241 VFKGDDVFNTHHILNLSLSKSKAKDLFGDVYHNGAPHRDKKLYIPEVFTGYV 300
 DB 241 VFKGDDVFNTHHILNLSLSKSKAKDLFGDVYHNGAPHRDKKLYIPEVFTGYV 300

QY 301 PPYAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKLGVLVPEKHKGFTDIEE 360
 DB 301 PPYAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKLGVLVPEKHKGFTDIEE 360

QY 361 KNKKNICSVIDMLVHSRKPQEMIDWSQLSPNLK 397
 DB 361 KNKKNICSVIDMLVHSRKPQEMIDWSQLSPNLK 397

RESULT 3
 Q8UWMO PRELIMINARY; PRT: 406 AA.

ID Q8UWMO
 AC Q8UWMO;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Beta-3-galactosyltransferase.
 GN SSP5.
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chou C.-M., Leu J.-H., Huang C.-J.;
 RT "A family of novel genes encoding beta-3-galactosyltransferase from
 zebrafish, zssp5 mRNA.";

```
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF321831; AAL32299.1; -.
DR InterPro: IPR002659; Glyco_trans_31.
DR Pfam: PF01762; Galactosyl_T; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 406 AA; 46953 MW; F928471BBDF1681 CRC64;

Query Match 57.8%; Score 1229.5; DB 13; Length 406;
Best Local Similarity 57.9%; Pred. No. 1.8e-104;
Matches 226; Conservative 63; Mismatches 98; Indels 3; Gaps 2;

Qy 5 RRRVKLLGLIMANVYIYLIVEVSKNSODKNGKGGVILPKFKWKPSTPRAYNREQE 64
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 5 RRVKVKMA--MMTWVFLFIVVEVSRNAGKSSKNNKSLVPLKRFWAKDLPSSDAYNRQQQ 62
Qy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 65 KLRWYNPILNRVANOTGELATSPN-TSHLSYCEPDSTVMTAVTDFNNLPDRFKDPLLYL 123
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 63 QINYNINRLEKYNIDNLPWLNDTVSLDSDPDYRVTTVKDINSLPDRFKDPLLYM 122
Qy 124 RCNRYSLLDIDOPKKCAKPFLLAIAKSLIPHFAARRQAIRESWGRETNGVTVRVFLLG 183
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 123 RCRSYPIVDQPNICKQKPFLLAIAKSLVPHFDRQAIRESWGKVGRIANRSVTVFLLG 182
Qy 184 KTPPEONHDPDLSMDLKFESDKHODILMWNRYDTFFNLSLKEVLFRLWVSTPCDAEFVFK 243
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 183 NAATEDHFPDLSKMLHSESIHRDILQWYRDTFFNLTKEVLFLEWLSLTCRPGANIFK 242
Qy 244 GDDVDFVNTTHILNLSKSKAKOLFIDGVTHNAGPHRDKKLYYIPEVFTGVPYPPY 303
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 243 GDDVDFVNTTHIDFLNLSNAKARELFVGDVITNAGPHRDKKLYYIPESFVGVMPAY 302
Qy 304 AGGGGFLYSPALLRYSATSRVHLYPIDVYTGMCQLKGLVPEKHKGFRFTDIEENK 363
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 303 AGGGGFLFSQLAQRUNISKLPLYPIDVYTGMCMLKMWGLAPEKHKGFRFTDIEEYR 362
Qy 364 KNICSYDILMLVHSRKPQEMIDIWSQLSQSP 393
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 363 DNACAYKSLMLVHPRSPQHMKIWMLNDP 392

RESULT 4
Q8UWM3 PRELIMINARY; PRT; 412 AA.
AC Q8UWM3
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DI 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Beta-3-galactosyltransferase.
GN SSP2.
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Chou C.-M., Leu J.-H., Huang C.-J.;
RT *A family of novel genes encoding beta-3-galactosyltransferase from
RT zebrafish, zssp2 mRNA.
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF321829; AAL32299.1; -.
DR InterPro: IPR002659; Glyco_trans_31.
DR Pfam: PF01762; Galactosyl_T; 1.
DR Transferase; Glycosyltransferase.
KW SSP2.
SQ SEQUENCE 412 AA; 47812 MW; 80C554675A2CB00A CRC64;

Query Match 32.3%; Score 688; DB 13; Length 412;
Best Local Similarity 37.2%; Pred. No. 1.1e-54;
Matches 135; Conservative 76; Mismatches 124; Indels 28; Gaps 5;

Qy 60 NREOEKLNLR-WYNPILNRVANOTGELATSPN-TSHLSYCEPDSTVMTAVTDFNNLPD 114
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 60 NREOEKLDIYIPLVN-----ASELPHPFRGIPRWLNHS-CKHDVNVTAQIEDFNLSLP 113
Qy 115 RFKDFLLYLCRNYSLLIDOPKKCAKPFLLAIAKSLIPHFAARRQAIRESWGRETNGVNG 174
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 114 RFQDFLLYMGCRSVPLITRAPKVCSPPPYLLAIAKSLAPHFDRQAIRESWGAGILDGO 173
Qy 175 TVRVFLLGKTPEDNHPDLSMDLKFESDKHODILMWNRYDTFFNLSLKEVLFRLWVST 234
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 174 RIATVFVLGNTASTDHPDLSNMVXKHEALYGDVQLQWYRDTFFNLTKEVLFLEWFGSH 233
Qy 235 CPDAEFVFKGDDVFNTHILNLSKSKAKOLFIDGVTHNAGPHRDKKLYYIPEV 294
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 234 CASAQYVFKGDDVFNTHILNLSKSKAKOLFIDGVTHNAGPHRDKKLYYIPEV 293
Qy 295 FYTGVPYPPYAGGGFLYSPALLRYSATSRVHLYPIDVYTGMCQLKGLVPEKHKGFR 354
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 294 VFTGGYPPYAGGGFLYSGVGLRLKISRLVLYPIDVYTGMCQLQRLGLVPEKHSGFK 353
Qy 355 TFDTEENKKNICSYDILMLVHSRKPQEMIDIWSQLSQSPNLC 397
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 354 TFDIEAKHRENCAYKGLDILVHPRSPQHMKIWMLNDPNAIC 396

RESULT 5
Q8UWM2 PRELIMINARY; PRT; 412 AA.
AC Q8UWM2
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DI 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Beta-3-galactosyltransferase.
GN SSP3.
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Chou C.-M., Leu J.-H., Huang C.-J.;
RT *A family of novel genes encoding beta-3-galactosyltransferase from
RT zebrafish, zssp2 mRNA.
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF321829; AAL32299.1; -.
DR InterPro: IPR002659; Glyco_trans_31.
DR Pfam: PF01762; Galactosyl_T; 1.
DR Transferase; Glycosyltransferase.
KW SSP3.
SQ SEQUENCE 412 AA; 47812 MW; 80C554675A2CB00A CRC64;

Query Match 32.3%; Score 688; DB 13; Length 412;
Best Local Similarity 37.2%; Pred. No. 1.1e-54;
Matches 135; Conservative 76; Mismatches 124; Indels 28; Gaps 5;

Qy 60 NREOEKLNLR-WYNPILNRVANOTGELATSPN-TSHLSYCEPDSTVMTAVTDFNNLPD 99
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 47 DKYVEVGRMCGP--QCPSPFRKNLKAVENTSSGSGSDSKRAFKPLPKKWDVNTCTEN 104
Qy 100 STVMTAVTDFNNLPDRFKDILYLCRNYSLLIDOPKKCAKPFLLAIAKSLIPHFAARR 159
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 105 STIKTQLW-FRRLSPRFHEPVLHRCRYFPMLLNHPKCGGGVDVLVVYVKSVEIHDRRE 163
Qy 160 AIRESWGRETNGVNGQTVRVFLLGKTPEDNHPDLSMDLKFESDKHODILMWNRYDTFFN 219
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 164 AVRTWKGQEIQGLKIKTILGLTGPAGKDSRNLAQLVQYEDRTYGDILQWDFMDTFEN 223
Qy 220 LSLKEVLFRLWVSTSCPDAEVFKGDDVFNTHILNLSKSKAKOLFIDGVTHNAGPHRDK 278
Dy |||||: || |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Dy 224 LTLKEVNLFWFSYICPDVFIKGGDDVFPVHTKNLVELIGFKENKENVNLIVDAILE 283
Qy 279 AGPHRDKKLYYIPEVFTGVPYPPYAGGGFLYSGPALLRYSATSRVHLYPIDVYTG 338
```



```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Leu J.H., Chou C.M., Huang C.J.:
RT "Cloning and expression of a novel human beta-1,3-
RL galactosyltransferase-related gene.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF321825; AAL37219.1; -.
DR InterPro; IPR002659; Glyco_trans_31.
DR Pfam; PF01762; Galactosyl_T; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 353 AA; 39847 MW; FB91097D5120AC50 CRC64;

Query Match 30.4%; Score 647.5; DB 4; Length 353;
Best Local Similarity 39.7%; Pred. No. 4.7e-51;
Matches 147; Conservative 53; Mismatches 115; Indels 55; Gaps 11;

QY 40 GVIIPEKFKWPPSTPRA---YNNREQEKLNRWYNPILNRVANQTGELATSPNTHLSYC 96
Db 19 GCLLFVRKTAAPAGDPTAHQPFW-----APTPRHRSRC 51
QY 97 EPDSTVMTAVTDNNLPDRFKDFLLYLRCRNYSLIDQPKKCAKPFLLLAIKSLIPHEA 156
Db 52 PPNHTVSSASL---SLPSRHLFLTYRHCNFSILL-EPSCGSKDTFLLAIAKSPGHVE 107
QY 157 RQAIRESWGRNVTG---NQTVVRFVLL---GKTPPEDNHPDLSDMLKFFESDKHQDIL 209
Db 108 RRAAIRSTWGR---VGGWARGROLKLVFLGVAGSAPP-----AQLLAYESREFDDIL 157
QY 210 MNRYRTFFNLISLKEVLFRLWVSTSCPDABFVFKGDDVFNTHILNLSLSKSKAKD 269
Db 158 QWDETFEFLNLTKELHLQRWVVAACPAHFMLKGGDDVFVHPVNVLEFLDGN--DPAQD 215
QY 270 LFGDVTNHNAGPRHKKLYIPEVFTGV-YPPYAGGGGFLYSGPALLRLYSATSRVHL 328
Db 216 LLVGDVIRQALPNRNTKVYFIPPSMYRATHYPPYAGGGGYVMSRATVRRLOAIMEDAEL 275
QY 329 YPIDDVYTGMLQKLGVLPEKHGKFRTEDEEK-NKNKICSYIDLMLVHSRKPQEMIDIW 387
Db 276 FPIDDVFGVCHLRGLSPMHAGFKTFGIRRLDPLDPCLYRGLLLVHRLSPLEMTW 335
QY 388 SLOQSPNLKC 397
Db 336 ALVTDEGLKC 345

RESULT 14
Q8WWR5 ID Q8WWR5 PRELIMINARY; PRT; 353 AA.
AC Q8WWR5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Beta 1,3-galactosyltransferase.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bennett E.P.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ278962; CAC82375.1; -.
DR InterPro; IPR002659; Glyco_trans_31.
DR Pfam; PF01762; Galactosyl_T; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 353 AA; 39863 MW; EAA7A35019C36D64 CRC64;

Query Match 30.4%; Score 647.5; DB 4; Length 353;
Best Local Similarity 39.7%; Pred. No. 4.7e-51;
Matches 147; Conservative 53; Mismatches 115; Indels 55; Gaps 11;

QY 40 GVIIPEKFKWPPSTPRA---YNNREQEKLNRWYNPILNRVANQTGELATSPNTHLSYC 96
Db 19 GCLLFVRKTAAPAGDPTAHQPFW-----APTPRHRSRC 51
QY 97 EPDSTVMTAVTDNNLPDRFKDFLLYLRCRNYSLIDQPKKCAKPFLLLAIKSLIPHEA 156
Db 52 PPNHTVSSASL---SLPSRHLFLTYRHCNFSILL-EPSCGSKDTFLLAIAKSPGHVE 107
QY 157 RQAIRESWGRNVTG---NQTVVRFVLL---GKTPPEDNHPDLSDMLKFFESDKHQDIL 209
Db 108 RRAAIRSTWGR---VGGWARGROLKLVFLGVAGSAPP-----AQLLAYESREFDDIL 157
QY 210 MNRYRTFFNLISLKEVLFRLWVSTSCPDABFVFKGDDVFNTHILNLSLSKSKAKD 269
Db 158 QWDETFEFLNLTKELHLQRWVVAACPAHFMLKGGDDVFVHPVNVLEFLDGN--DPAQD 215
QY 270 LFGDVTNHNAGPRHKKLYIPEVFTGV-YPPYAGGGGFLYSGPALLRLYSATSRVHL 328
Db 216 LLVGDVIRQALPNRNTKVYFIPPSMYRATHYPPYAGGGGYVMSRATVRRLOAIMEDAEL 275
QY 329 YPIDDVYTGMLQKLGVLPEKHGKFRTEDEEK-NKNKICSYIDLMLVHSRKPQEMIDIW 387
Db 276 FPIDDVFGVCHLRGLSPMHAGFKTFGIRRLDPLDPCLYRGLLLVHRLSPLEMTW 335
QY 388 SLOQSPNLKC 397
Db 336 ALVTDEGLKC 345

RESULT 14
Q8WWR5 ID Q8WWR5 PRELIMINARY; PRT; 378 AA.
AC Q8WWR5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Beta-1,3-N-acetylglucosaminyltransferase BGN-T4.
GN BGN-T4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shiraishi N., Natsume A., Toyayachi A., Endo T., Akashima T.,
RA Yamada Y., Imai N., Nakagawa S., Koizumi S., Sekine S., Narimatsu H.,
RA Sasaki K.;
RL J. Biol. Chem. 276:3498-3507(2001).
DR EMBL; AB049586; BAB21532.1; -.
DR Genew; HGNC:15683; B3GNT4.
DR InterPro; IPR002659; Glyco_trans_31.
DR Pfam; PF01762; Galactosyl_T; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 378 AA; 42310 MW; E8E88665A36057E CRC64;

Query Match 30.4%; Score 647.5; DB 4; Length 378;
Best Local Similarity 39.7%; Pred. No. 5.2e-51;
Matches 147; Conservative 53; Mismatches 115; Indels 55; Gaps 11;

QY 40 GVIIPEKFKWPPSTPRA---YNNREQEKLNRWYNPILNRVANQTGELATSPNTHLSYC 96
Db 44 GCLLFVRKTAAPAGDPTAHQPFW-----APTPRHRSRC 76
QY 97 EPDSTVMTAVTDNNLPDRFKDFLLYLRCRNYSLIDQPKKCAKPFLLLAIKSLIPHEA 156
Db 77 PPNHTVSSASL---SLPSRHLFLTYRHCNFSILL-EPSCGSKDTFLLAIAKSPGHVE 132
QY 157 RQAIRESWGRNVTG---NQTVVRFVLL---GKTPPEDNHPDLSDMLKFFESDKHQDIL 209
Db 133 RRAAIRSTWGR---VGGWARGROLKLVFLGVAGSAPP-----AQLLAYESREFDDIL 182

```

```

Qy 210 MMYRDFEENLSKEVLFGRVSTSGPDAEFVKGGDDVFVNTHTILNYLNSLSKSKAKD 269
Db 183 QWDFTEDFNLTLKELHLQRWVVAACQAHFMUKGGDDVFVHPNVLEFDGW--DPAQD 240
Qy 270 LFIGDVHNGAPHRDKKLYYIPEVFTGV-YPPYAGGGGGLYSGPALLRLYSATSRVHL 328
Db 241 LLVGDVIRQALPNRNTKVKYFIPPSMYRATHYPPYAGGGGYVMSRATVRRLOAIMEDAE 300
Qy 329 YPIDDVYTGMLQKLGSLVPEKHKGFRFTDIEEK-NKKNICSYIDLMLVHSRKPQEMIDIW 387
Db 301 FPIDDDVFGMLRLRLGLSPMHAGFKTFGIRRLPLDPLDCLYRGLLLVHRLSPLEMTW 360
Qy 388 SQLQSPNLKC 397
Db 361 ALVTDEGLKC 370

```

Search completed: October 20, 2003, 23:04:02
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 22:49:49 ; Search time 38.5 Seconds
(without alignments)
1636.739 Million cell updates/sec

Title: US-09-804-357b-14

Perfect score: 2123

Sequence: 1 MSVGRRIKLLGLMANVF.....RKPEMIDWSOLQSAHLAC 397

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2123	100.0	397	19	AAW80212 Human cardiac and
2	2123	100.0	397	21	AAAY84683 Amino acid sequenc
3	2123	100.0	397	22	AAG66118 Human LiG46 polype
4	2123	100.0	397	23	AAE29093 Human LiG46 protei
5	2123	100.0	397	24	ABB82649 Human beta3gnt pol
6	2105	99.2	413	22	ABA49748 Human beta 1,3-N-a
7	2022	95.2	397	21	AAAY69698 Human brainiac pro
8	2007	94.5	373	22	AAB93536 Human protein sequ
9	1872	88.2	397	21	AAAY84641 Amino acid sequenc

10	1872	88.2	397	21	AAAY79953	Murine brainiac pr
11	1872	88.2	397	21	AAAY69697	Murine Brainiac pr
12	1872	88.2	397	22	AAG66115	Murine LiG46 polyp
13	1872	88.2	397	23	AAE29092	Murine LiG46 prote
14	1866	87.9	397	24	ABB82648	Murine beta3gnt po
15	1823	85.9	397	21	AAAY79954	Human brainiac pro
16	1743	82.1	367	22	AAG66116	Murine LiG46 matur
17	1068	50.3	197	23	AAE15936	Human 7023630/124-
18	667	31.4	401	23	AAE22148	Human TRNFR-10 pro
19	642.5	30.3	364	23	ABU65191	Human NOV106a pro
20	641.5	30.2	334	22	ABG20251	Novel human diagno
21	634.5	29.9	415	23	ABG30984	Human galactosyltr
22	632	29.8	352	20	AAAY06462	Human Brainiac-3.
23	632	29.8	352	24	ABR47731	Human secreted pro
24	632	29.8	353	20	AAAY36224	Human secreted pro
25	628.5	29.6	378	21	AAAB24033	Human PRO4344 prot
26	628.5	29.6	378	22	AAAB49751	Human beta 1,3-N-a
27	628.5	29.6	378	23	ABG34047	Human pro peptide
28	616.5	29.0	378	21	AAAB07435	The beta-1,3-galac
29	616.5	29.0	378	23	ABG32941	Human galactosyltr
30	616.5	29.0	378	23	AAE24688	Human znssp6 prote
31	614.5	28.9	384	23	ABG32374	Novel human enzyme
32	602	28.4	402	21	AAAB24035	Human PRO4397 prot
33	602	28.4	402	23	AAE22142	Human TRNFR-4 prot
34	602	28.4	402	23	ABG34046	Human pro peptide
35	601	28.3	372	19	AAW64558	Human epidermoid c
36	601	28.3	372	22	AAU29167	Human p80 polypept
37	601	28.3	372	22	AAAB49749	Human beta 1,3-N-a
38	601	28.3	372	23	ABBO9716	Amino acid sequenc
39	601	28.3	372	23	AAU11272	Human beta1,3-N-ac
40	601	28.3	372	24	ABU71255	Human PRO1266 prot
41	601	28.3	372	24	ABU65712	Human secreted/tri
42	601	28.3	372	24	ABU66045	Novel human secret
43	601	28.3	372	24	ABU67549	Human secreted/tri
44	601	28.3	372	24	ABR01801	Human cancer-relat
45	601	28.3	372	24	ABU65407	human PRO polypept

ALIGNMENTS

RESULT 1

AAW80212
ID AAW80212 standard; Protein; 397 AA.

XX AAW80212;

XX AC

XX AC

DI 18-JAN-1999 (first entry)

XX Human cardiac and pancreatic protein (CAPP).

XX Cardiac and pancreatic protein; CAPP; muscle-derived growth factor;

KW human; pancreatitis; myocardial infarction; cardiomyopathy.

XX Homo sapiens.

OS Homo sapiens.

XX Key

XX Location/Qualifiers

XX Peptide

XX 1..32

XX /label= sig_peptide

XX 33..397

XX /label= Mat_protein

XX 1..11

XX /note= "epitope-bearing peptide (Claim 11)"

XX 29..72

XX /note= "epitope-bearing peptide (Claim 11)"

XX 78..89

XX /note= "epitope-bearing peptide (Claim 11)"

XX 94..105

XX /note= "epitope-bearing peptide (Claim 11)"

XX 110..119

XX /note= "epitope-bearing peptide (Claim 11)"

XX 124..142

XX /note= "epitope-bearing peptide (Claim 11)"

XX /note= "epitope-bearing peptide (Claim 11)"

```
FT Peptide 151..176
FT /note= "epitope-bearing peptide (Claim 11)"
FT Peptide 184..218
FT /note= "epitope-bearing peptide (Claim 11)"
FT Peptide 232..251
FT /note= "epitope-bearing peptide (Claim 11)"
FT Peptide 262..272
FT /note= "epitope-bearing peptide (Claim 11)"
FT Peptide 280..290
FT /note= "epitope-bearing peptide (Claim 11)"
FT Peptide 346..368
FT /note= "epitope-bearing peptide (Claim 11)"
FT Peptide 376..385
FT /note= "epitope-bearing peptide (Claim 11)"
XX
XX WO9844112-A1.
XX
XX 08-OCT-1998.
XX
XX 27-MAR-1998; 98WO-US06022.
XX
XX 28-MAR-1997; 97US-0042855.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Soppet DR;
XX
XX WPI: 1998-557111/47.
XX N-PSDB; AAV66367.
XX
XX New isolated cardiac and pancreatic protein - used to develop
XX products for the diagnosis and treatment of e.g. pancreatitis or
XX abnormal hypertrophy of the heart
XX
XX Claim 1: Fig 1: 113pp; English.
XX
XX This is the amino acid sequence of novel cardiac and pancreatic
XX protein (CAPP), a member of the muscle-derived growth factor
XX superfamily, as deduced from the nucleotide sequence (see AAV66367)
XX of a cDNA clone discovered in a cDNA library derived from activated
XX T cells. The gene is abundant in adult heart and pancreas. CAPP
XX isolated nucleic acid molecules and polypeptides are provided, as
XX are vectors, host cells and recombinant methods for producing the
XX same. The invention further relates to screening methods for
XX identifying agonists and antagonists of CAPP activity, as well as
XX diagnostic methods for detecting the presence of activated T-cells,
XX and mature heart, pancreas and placental tissues and cells. The
XX CAPP polypeptides can modulate the differentiation and proliferation
XX of cells and tissue, both in vivo and ex vivo. The products can be
XX used in the diagnosis and treatment of pancreatitis and conditions
XX that cause abnormal hypertrophy of the heart, such as hypertension,
XX myocardial infarction, valve disease and cardiomyopathy. The
XX products can also be used in detection and cell culturing.
XX
XX Sequence 397 AA;
XX
XX Query Match 100.0%; Score 2123; DB 19; Length 397;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-218;
XX Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 MSVGRRIKILGLILMANVFIYIMEVSKSSOKEKNGKEVIIPKPKFKWIKSTPPPAYWN 60
XX |
XX 1 MSVGRRIKILGLILMANVFIYIMEVSKSSOKEKNGKEVIIPKPKFKWIKSTPPPAYWN 60
XX |
XX 61 REQEKLNRQNPILSMLTNTGTAGRLSNLSHLYNCEPDLRVTSVVTGNNLPDRFKDEL 120
XX |
XX 61 REQEKLNRQNPILSMLTNTGTAGRLSNLSHLYNCEPDLRVTSVVTGNNLPDRFKDEL 120
XX |
XX 121 LYLRCRNYSLIDOPDKCAKKPFLLLAIAKSLTFHFARROAIRSWQESNAGNQTVRVVF 180
XX |
XX 121 LYLRCRNYSLIDOPDKCAKKPFLLLAIAKSLTFHFARROAIRSWQESNAGNQTVRVVF 180
XX |
XX 181 LLGOTPPEDNHPDLSMDLKFESEKHODILMNNYRDTFFNLSLKEVFLRWVSTSCPDTEF 240
XX |
DB Db 181 LLGOTPPEDNHPDLSMDLKFESEKHODILMNNYRDTFFNLSLKEVFLRWVSTSCPDTEF 240
Qy Qy 241 VFKGDDDDVFVNTTHILNLSLSKTRAKOLFIGNVHNAGPHRDKLKYIPEVYVSGLY 300
Db Db 241 VFKGDDDDVFVNTTHILNLSLSKTRAKOLFIGNVHNAGPHRDKLKYIPEVYVSGLY 300
Qy Qy 301 PPYAGGGGFLYSGLHALRLYHITDOVHLYPIDOVYTGMCLOKLGLYPEKHKGRTFDIEE 360
Db Db 301 PPYAGGGGFLYSGLHALRLYHITDOVHLYPIDOVYTGMCLOKLGLYPEKHKGRTFDIEE 360
Qy Qy 361 KNKNKICSYVDLMLVHSRKPQEMIDWSQSAHLKC 397
Db Db 361 KNKNKICSYVDLMLVHSRKPQEMIDWSQSAHLKC 397
XX
XX RESULT 2
XX AAY84683
XX ID AAY84683 standard; Protein: 397 AA.
XX
XX AC AAY84683;
XX
XX DT 25-JUL-2000 (first entry)
XX
XX DE Amino acid sequence of a human LIG46 polypeptide.
XX
XX KW Leptin; LIG46; body weight; leptin inducible gene; obesity; cachexia;
XX KW LIG56; Tgtp; LRG-47; RC10-II; Stral3.
XX
XX OS Homo sapiens.
XX
XX PN WO200015826-A2.
XX
XX PD 23-MAR-2000.
XX
XX PF 10-SEP-1999; 99WO-US20722.
XX
XX PR 10-SEP-1998; 98US-0150857.
XX PR 29-OCT-1998; 98US-0106378.
XX PR 19-NOV-1998; 98US-0195896.
XX PR 15-APR-1999; 99US-0292228.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX White D, Zhou J, Tartaglia LA;
XX
XX WPI: 2000-271461/23.
XX N-PSDB; AAA12716.
XX
XX Method for determining compounds which modulate body weight and can be
XX used to treat e.g. obesity comprises measuring the activity of leptin
XX inducible genes.
XX
XX Disclosure; Fig 8; 123pp; English.
XX
XX The present sequence represents a leptin induced LIG46 polypeptide. The
XX specification describes a method for determining whether a compound can
XX be used to modulate body weight by measuring the activity of leptin
XX inducible genes, such as LIG46. The method can be used to specifically
XX detect LIG46 and LIG56 nucleic acid molecules relative to other nucleic
XX acid molecules encoding galactosyltransferases or GTP-binding proteins.
XX They can also be used in diagnostic assays to identify the presence or
XX absence of a genetic lesion or mutation characterized by aberrant
XX translational modification of the genes or aberrant post-
XX and nucleic acid molecules can be used to treat obesity or cachexia.
XX LIG46 and LIG56 antagonists are used to treat obesity and LIG46 and
XX LIG56 agonists are used to treat low body weight. The leptin inducible
XX genes i.e. LIG46, LIG56, Tgtp, LRG-47, RC10-II and Stral3 can be used
XX to determine whether a compound modulates body weight and can then be
XX used to treat obesity or cachexia or low body weight.
XX
XX Sequence 397 AA;
XX SQ
```

```
Query Match      100.0%; Score 2123; DB 21; Length 397;
Best Local Similarity 100.0%; Pred. No. 1.4e-218;
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSVGRRIKLLGILMANVFIYFIMEVSKSSQKNGKEVIIPKFKWKISTPPEAYWN 60
DB 1 MSVGRRIKLLGILMANVFIYFIMEVSKSSQKNGKEVIIPKFKWKISTPPEAYWN 60
QY 61 REQKLNQYNPILSMLTNOTGAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRFKDFL 120
DB 61 REQKLNQYNPILSMLTNOTGAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRFKDFL 120
QY 121 LYLCRNYSLLIDQPKCAKPPFLLLAISKSLTPHFARRQAIRESWGQESNAGNQTIVRVF 180
DB 121 LYLCRNYSLLIDQPKCAKPPFLLLAISKSLTPHFARRQAIRESWGQESNAGNQTIVRVF 180
QY 181 LLGQTPPEDNHPDLSMDLKFESEKHQDILMNNYRTDFFNLSKEVFLRWVSTSCPDTEF 240
DB 181 LLGQTPPEDNHPDLSMDLKFESEKHQDILMNNYRTDFFNLSKEVFLRWVSTSCPDTEF 240
QY 241 VFKGDDVFNTHILNLSKTKAKDLFIGDVIHNAIPHRRDKKLYIPEVYVSGLY 300
DB 241 VFKGDDVFNTHILNLSKTKAKDLFIGDVIHNAIPHRRDKKLYIPEVYVSGLY 300
QY 301 PPYAGGGFLYSGHLALRLYHITDQVHLYPIDDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360
DB 301 PPYAGGGFLYSGHLALRLYHITDQVHLYPIDDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360
QY 361 KNNKNICSYVDLMLVHSRKPQEMIDWSQLSAHLKC 397
DB 361 KNNKNICSYVDLMLVHSRKPQEMIDWSQLSAHLKC 397

RESULT 3
ID AAG66118 standard; Protein; 397 AA.
AC AAG66118;
XX 13-MAR-2002 (first entry)
XX Human LIG46 polypeptide.
XX Leptin; LIG46; LIG56; leptin induced gene; Tgtp; LRG-47; RC10-II; Scral3;
XX anorectic; anabolic; antisense therapy; human.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Peptide 1..32
XX Protein /note= "signal peptide"
XX /note= "mature protein"
XX US2001024808-A1.
XX 27-SEP-2001.
XX 12-MAR-2001; 2001US-0804357.
XX 29-OCT-1998; 98US-106378P.
XX 19-NOV-1998; 98US-0195896.
XX 10-SEP-1998; 98US-0150857.
XX (MILL-) MILLENNIUM PHARM INC.
XX White D, Zhou J, Tartaglia LA;
XX WPI: 2001-624963/72.
XX DR N-PSDB; AA167869.
XX An isolated polypeptide useful for treating a weight disorder and for
```

```
PT screening compounds that may inhibit protein activity comprises a
PT leptin-induced protein
XX Example 2; Fig 8; 46pp; English.
XX The invention relates to genes whose expression are induced by leptin.
XX LIG46 and LIG56 are novel leptin induced genes (LIG), while four other
XX genes such as Tgtp, LRG-47, RC10-II and Scral3 have been previously
XX identified. The leptin induced proteins can be expressed by standard
XX recombinant methodology. The proteins and encoding polynucleotides may
XX be used in screening assays to identify compounds that may bind to it.
XX Administering a molecule, e.g., an antisense molecule, which reduces
XX expression of activity of protein selected from LIG46, LIG56, Tgtp,
XX LRG-47, RC10-II, and Scral3 may be used to treat a weight disorder. The
XX method may further comprise administering leptin. The present sequence
XX represents a human LIG46 protein.
SQ Sequence 397 AA;
Query Match      100.0%; Score 2123; DB 22; Length 397;
Best Local Similarity 100.0%; Pred. No. 1.4e-218;
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSVGRRIKLLGILMANVFIYFIMEVSKSSQKNGKEVIIPKFKWKISTPPEAYWN 60
DB 1 MSVGRRIKLLGILMANVFIYFIMEVSKSSQKNGKEVIIPKFKWKISTPPEAYWN 60
QY 61 REQKLNQYNPILSMLTNOTGAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRFKDFL 120
DB 61 REQKLNQYNPILSMLTNOTGAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRFKDFL 120
QY 121 LYLCRNYSLLIDQPKCAKPPFLLLAISKSLTPHFARRQAIRESWGQESNAGNQTIVRVF 180
DB 121 LYLCRNYSLLIDQPKCAKPPFLLLAISKSLTPHFARRQAIRESWGQESNAGNQTIVRVF 180
QY 181 LLGQTPPEDNHPDLSMDLKFESEKHQDILMNNYRTDFFNLSKEVFLRWVSTSCPDTEF 240
DB 181 LLGQTPPEDNHPDLSMDLKFESEKHQDILMNNYRTDFFNLSKEVFLRWVSTSCPDTEF 240
QY 241 VFKGDDVFNTHILNLSKTKAKDLFIGDVIHNAIPHRRDKKLYIPEVYVSGLY 300
DB 241 VFKGDDVFNTHILNLSKTKAKDLFIGDVIHNAIPHRRDKKLYIPEVYVSGLY 300
QY 301 PPYAGGGFLYSGHLALRLYHITDQVHLYPIDDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360
DB 301 PPYAGGGFLYSGHLALRLYHITDQVHLYPIDDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360
QY 361 KNNKNICSYVDLMLVHSRKPQEMIDWSQLSAHLKC 397
DB 361 KNNKNICSYVDLMLVHSRKPQEMIDWSQLSAHLKC 397

RESULT 4
AAE29093
ID AAE29093 standard; Protein; 397 AA.
XX AAE29093;
XX 27-JAN-2003 (first entry)
XX Human LIG46 protein.
XX LIG46; leptin; low body weight; chromosomal mapping; tissue typing;
XX forensic biology; transgenic; gene therapy; antianorectic; human;
XX chromosome 2.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Peptide 1..32
XX Protein /label= "signal_peptide"
XX /note= "Human mature LIG46 protein"
```

XX WO200274905-A2.
XX 26-SEP-2002.
XX 20-NOV-2001; 2001WO-USA3345.
XX 21-NOV-2000; 2000US-0717778.
XX (MILL-) MILLENNIUM PHARM INC.
XX White DW, Zhou J, Tartaglia LA, Stricker-krongrad A, Clausen H;
XX WPI: 2002-759886/82.
XX N-PSDB; AAD46673.
XX New isolated nucleic acid molecules, designated as leptin-induced genes
XX 46 (LIG46), useful for treating a subject having a disorder
XX characterized by undesirable level of LIG46 expression or activity,
XX such as low body weight -
XX Claim 22; Fig 6; 90pp; English.
XX The invention relates to LIG46, a gene whose expression is induced by
XX leptin. LIG46 DNA and protein are useful in treating a subject having
XX a disorder characterized by undesirable level of LIG46 expression or
XX activity, such as low body weight. They are also useful in a screening
XX assay, chromosomal mapping, tissue typing and forensic biology. The
XX probes based on the LIG46 nucleotide sequence are useful for detecting
XX transcripts or genomic sequences encoding the same or related proteins.
XX The LIG46 polypeptides are useful as immunogens for raising anti-LIG46
XX antibodies. The host cells are useful for producing non-human transgenic
XX animals. LIG46 DNA is used in gene therapy. The present sequence is
XX human LIG46 protein. LIG46 gene is located at chromosome 2.
XX
XX Sequence 397 AA;
Query Match 100.0%; Score 2123; DB 23; Length 397;
Best Local Similarity 100.0%; Pred. No. 1.4e-218;
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSVGRRIKLLGILMANVFIYFIMEVSKSSQKNGKEVVIIPKEFKWISTPPEAYWN 60
DB 1 MSVGRRIKLLGILMANVFIYFIMEVSKSSQKNGKEVVIIPKEFKWISTPPEAYWN 60
QY 61 REQEKLNQYNPILSMLTNOTGEAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRKDFL 120
DB 51 REQEKLNQYNPILSMLTNOTGEAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRKDFL 120
QY 121 LYLRCRNYSLIDQPKCAKKPFLLLAIAKSLTPHFARRQAIRESWGQESNAGNQTIVRVF 180
DB 121 LYLRCRNYSLIDQPKCAKKPFLLLAIAKSLTPHFARRQAIRESWGQESNAGNQTIVRVF 180
QY 181 LLGQTPPEDNHPDLSMDLKFESKQHDILMWNRTDFNLSLKEVFLRWVSTSCPDTEF 240
DB 181 LLGQTPPEDNHPDLSMDLKFESKQHDILMWNRTDFNLSLKEVFLRWVSTSCPDTEF 240
QY 241 VFKGDDDFVNTHTLNLNLSKTKAKOLFIDGVIHNAGPHRDKKLKYYIPVYVYSGLY 300
DB 241 VFKGDDDFVNTHTLNLNLSKTKAKOLFIDGVIHNAGPHRDKKLKYYIPVYVYSGLY 300
QY 301 PPYAGGGFLYSGHLARLTHITDQVHLYPDIDVYVYGMCLQKLGVLPEKHKGFRTEIDEE 360
DB 301 PPYAGGGFLYSGHLARLTHITDQVHLYPDIDVYVYGMCLQKLGVLPEKHKGFRTEIDEE 360
QY 361 KNKNKICSYVDLMLVHSRKPQEMIDVWSQSAHLKC 397
DB 361 KNKNKICSYVDLMLVHSRKPQEMIDVWSQSAHLKC 397
RESULT 5
ABB82649
ID ABB82649 standard; Protein; 397 AA.

XX ABB82649;
XX 19-FFB-2003 (first entry)
XX Human beta3Gnt polypeptide sequence.
XX Beta3Gnt; transgenic; beta-1,3-N-acetylglucosaminyltransferase;
XX disease system; human; enzyme.
XX Homo sapiens.
XX WO200279413-A2.
XX 10-OCT-2002.
XX 29-MAR-2002; 2002WO-US09645.
XX 29-MAR-2001; 2001US-280706P.
XX 28-MAR-2002; 2002US-0280706.
XX (DELT-) DELTAGEN INC.
XX Leviten MW, Phillips R;
XX WPI: 2003-067437/06.
XX N-PSDB; ABV75082.
XX New transgenic mouse comprising a disruption in a beta3Gnt gene, as in
XX vivo model to study various disease states or conditions in which
XX beta3Gnt may be implicated or involved, such as abnormal cell growth,
XX cancer and metastasis -
XX Disclosure; Fig 4; 55pp; English.
XX The invention relates to a transgenic mouse comprising a disruption in a
XX beta3Gnt (beta-1,3-N-acetylglucosaminyltransferase) gene, where there is
XX no native expression of beta3Gnt gene. The transgenic mice may be used
XX as in vivo model to study various disease states or conditions in which
XX beta3Gnt may be implicated or involved, such as abnormal cell growth,
XX cancer and metastasis, and to evaluate various treatments or to identify
XX agents for treating disease states or conditions, such as anxiety or
XX depression. Animal-based disease systems may be used to identify
XX compounds capable of ameliorating disease symptoms, as test substrates
XX for the identification of drugs, pharmaceuticals, therapies and
XX interventions that may be effective in treating a disease or other
XX phenotypic characteristic of the animal. The present sequence represents
XX a human beta3Gnt polypeptide sequence.
XX Sequence 397 AA;
Query Match 100.0%; Score 2123; DB 24; Length 397;
Best Local Similarity 100.0%; Pred. No. 1.4e-218;
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSVGRRIKLLGILMANVFIYFIMEVSKSSQKNGKEVVIIPKEFKWISTPPEAYWN 60
DB 1 MSVGRRIKLLGILMANVFIYFIMEVSKSSQKNGKEVVIIPKEFKWISTPPEAYWN 60
QY 61 REQEKLNQYNPILSMLTNOTGEAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRKDFL 120
DB 61 REQEKLNQYNPILSMLTNOTGEAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRKDFL 120
QY 121 LYLRCRNYSLIDQPKCAKKPFLLLAIAKSLTPHFARRQAIRESWGQESNAGNQTIVRVF 180
DB 121 LYLRCRNYSLIDQPKCAKKPFLLLAIAKSLTPHFARRQAIRESWGQESNAGNQTIVRVF 180
QY 181 LLGQTPPEDNHPDLSMDLKFESKQHDILMWNRTDFNLSLKEVFLRWVSTSCPDTEF 240
DB 181 LLGQTPPEDNHPDLSMDLKFESKQHDILMWNRTDFNLSLKEVFLRWVSTSCPDTEF 240
QY 241 VFKGDDDFVNTHTLNLNLSKTKAKOLFIDGVIHNAGPHRDKKLKYYIPVYVYSGLY 300
DB 241 VFKGDDDFVNTHTLNLNLSKTKAKOLFIDGVIHNAGPHRDKKLKYYIPVYVYSGLY 300

DD 241 VFKGDDOVFNTHILNLSLTKAKDLFIGDVHNNAGPHRDKLKYIPEVYVSGLY 300
 QY 301 PPVAGGGFLYSGHLALRLYHITDQVHLYPIDVYTGMLQKLGVLPEKHKGFRTDIEE 360
 DB 301 PPVAGGGFLYSGHLALRLYHITDQVHLYPIDVYTGMLQKLGVLPEKHKGFRTDIEE 360
 QY 361 KKNKNICSYVDLMLVHSRKPQEMIDWSQLQSAHLKC 397
 DB 361 KKNKNICSYVDLMLVHSRKPQEMIDWSQLQSAHLKC 397

RESULT 6
 AAB49748
 ID AAB49748 standard; protein: 413 AA.
 XX AC AAB49748;
 XX DT 17-APR-2001 (first entry)
 XX DE Human beta 1,3-N-acetylglucosamine transferase protein G3.
 XX KW Sugar chain synthesis agent; beta 1,3-N-acetylglucosamine transferase;
 XX KW inflammation; cancer; metastasis; human.
 XX OS Homo sapiens.
 XX PN WO200100848-Al.
 XX PD 04-JAN-2001.
 XX PF 29-JUN-2000; 2000WO-JP04304.
 XX PR 29-JUN-1999; 99JP-0183437.
 XX PR 16-MAR-2000; 2000JP-0074757.
 XX PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX PI Sasaki K, Shiraishi N, Natsume A, Yamada Y, Nakagawa S, Sekine S;
 XX WPI; 2001-102895/11.
 XX DR N-PSDB; AAF29255.
 XX PT New polypeptide having beta1,3-N-acetylglucosamine transferase activity
 PT for diagnosis of inflammation, cancer and cancer metastasis,
 PT development of remedies, and for producing glycoconjugates -
 XX PS Claim 1; Page 147-150; 195pp; Japanese.
 XX CC This invention relates to a sugar chain synthesizing agent that contains
 CC a polypeptide as the active ingredient, where the polypeptide has beta
 CC 1,3-N-acetylglucosamine transferase activity. The polypeptides, DNA
 CC sequences encoding them and antibodies directed against the proteins are
 CC useful in the diagnosis of inflammation, cancer and its metastasis,
 CC development of remedies, and for producing sugar chains and
 CC glycoconjugates. Sequences AAB49748 - AAB49751 represent polypeptides of
 CC the invention, having beta 1,3-N-acetylglucosamine transferase activity,
 CC and polynucleotides AAF29255 - AAF29258 represent cDNA encoding them. PCR
 CC primers used in the isolation and characterisation of the cDNA sequences
 CC are represented by sequences ANF29259 - AAF29290.

SQ Sequence 413 AA;
 Query Match 99.2%; Score 2105; DB 22; Length 413;
 Best Local Similarity 96.1%; Pred. No. 1.3e-216;
 Matches 397; Conservative 0; Mismatches 0; Indels 16; Gaps 1;

QY 1 MSVGRRRKILGLM-----ANVFIFIMEVSKSSQKNGKGEVIIP 44
 DB 1 MSVGRRRKILGLMANNVFIFIMEVSKSSANVFIFIMEVSKSSQKNGKGEVIIP 60
 QY 45 KEKFWKISTPPEAYNREQELNRYNPILSMLTNOTGEAGRLSNISHLNCEPDLRVTS 104
 DB 61 KEKFWKISTPPEAYNREQELNRYNPILSMLTNOTGEAGRLSNISHLNCEPDLRVTS 120

QY 105 VVTGFNNLPDRFKDFLLYLCRCRNYSLIDQPKCAKPKFLLLAITSUTPHFARROAIRES 164
 DB 121 VVTGFNNLPDRFKDFLLYLCRCRNYSLIDQPKCAKPKFLLLAITSUTPHFARROAIRES 180
 QY 165 WQESNAGNQTIVRVFLLGQTTPEDNHPDLSMDLKFESEKHQDILMWNKYDTEFNLSLKE 224
 DB 181 WQESNAGNQTIVRVFLLGQTTPEDNHPDLSMDLKFESEKHQDILMWNKYDTEFNLSLKE 240
 QY 225 VLFLRWYSTSCPDTFVFVKGDDVFNTHILNLSLTKAKDLFIGDVHNNAGPHRD 284
 DB 241 VLFLRWYSTSCPDTFVFVKGDDVFNTHILNLSLTKAKDLFIGDVHNNAGPHRD 300
 QY 285 KKLKYIPEVYVSGLYPPYAGGGFLYSGHLALRLYHITDQVHLYPIDVYTGMLQKLG 344
 DB 301 KKLKYIPEVYVSGLYPPYAGGGFLYSGHLALRLYHITDQVHLYPIDVYTGMLQKLG 360
 QY 345 LVPEKHKGFRTDIEEKNKNICSYVDLMLVHSRKPQEMIDWSQLQSAHLKC 397
 DB 361 LVPEKHKGFRTDIEEKNKNICSYVDLMLVHSRKPQEMIDWSQLQSAHLKC 413

RESULT 7
 AAV69698
 ID AAV69698 standard; Protein: 397 AA.
 XX AC AAV69698;
 XX DT 08-MAY-2000 (first entry)
 XX DE Human Brainiac protein.
 XX KW Brainiac; human; mammalian; expressed sequence tag; EST;
 KW Drosophila melanogaster; Egghead; Notch; epithelial cell adhesion;
 KW Fringe family; cell viability; growth regulation; cell fate; cancer;
 KW psoriasis; skin lesion; nervous system disorder; developmental syndrome;
 KW transgenic animal; drug screening.
 XX OS Homo sapiens.
 XX FT Key Location/Qualifiers
 FT Misc-difference 104 /note= "Encoded by NNNNNN"
 XX CA2255109-Al.
 XX PD 17-JUN-1999.
 XX PF 17-DEC-1998; 98CA-2255109.
 XX PR 17-DEC-1997; 97CA-2225126.
 XX PA (HSCR-) HSC RES & DEV LP.
 XX PI Egan SE;
 XX WPI; 2000-148082/14.
 XX DR N-PSDB; AAZ87186.
 XX PT New nucleic acids encoding a murine and human Brainiac protein, useful
 PT for detecting somatic or germline DNA lesions which are responsible for
 PT developmental syndromes or diseases including cancer -
 XX PS Claim 13; Page31; 40pp; English.
 XX CC This sequence represents human Brainiac protein. The complete murine
 CC Brainiac cDNA sequence (AAZ87185) was used to screen an EST (expressed
 CC sequence tag) database to identify human Brainiac cDNA clones,
 CC which were assembled to form the human Brainiac cDNA of the present
 CC invention. Human Brainiac has significant similarity to Drosophila
 CC melanogaster Brainiac. Drosophila Brainiac and Egghead proteins regulate
 CC adhesion between epithelial cells, this activity being dependent on the
 CC presence of Notch. Drosophila Brainiac is a secreted protein, and has

CC sequence similarities with the Drosophila Fringe proteins. Brainiac/
CC Egghead-mediated epithelial cell adhesion is required for cell
CC viability, cell growth regulation and cell fate specification. Wild-type
CC or mutant forms of mammalian Brainiac proteins may therefore be used to
CC alter epithelial cell adhesion in a mammal. Mammalian Brainiac proteins,
CC active fragments analogues, and nucleic acids may be used to treat
CC diseases such as cancer, psoriasis and other skin lesions, and nervous
CC system disorders. Mammalian Brainiac nucleic acids may also be used to
CC detect somatic or germline DNA lesions which are responsible for
CC developmental syndromes or diseases including cancer. The mammalian
CC Brainiac proteins and fragments or its analogues are useful as antigens
CC in immunoassays including enzyme-linked immunosorbent assays (ELISA),
CC radioimmunoassays (RIA) and other non-enzyme linked antibody binding
CC assays. Non-human transgenic animals comprising nucleotide sequences
CC encoding human Brainiac protein (AA96969) can be used as animal models
CC for the study of mammalian Brainiac gene function, for the screening of
CC candidate compounds and for the evaluation of potential therapeutic
CC interventions.

XX Sequence 397 AA:

Query Match 95.2%; Score 2022; DB 21; Length 397;
Best Local Similarity 97.7%; Pred. No. 9.5e-208;
Matches 379; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 10 LIGILMANNVYIFINEVSKSSQKNGKGEVLIIPKEFWKISTPPEAYWNREQEKLNQ 69
DB 10 LIGILMANNVYIFINEVSKSSQKNGKGEVLIIPKEFWKISTPPEAYWNREQEKLNQ 69
QY 70 YNPILSMLTNGEAGRLSNISHLNCEPDLRVTSVVTGFNNLPDRFKDFLLYLRCRYS 129
DB 70 YNPILSMLTNGEAGRLSNISHLNCEPDLRVTSVVTGFNNLPDRFKDFLLYLRCRYS 129
QY 130 LLTDQPKCAKPFLLAIAKSLTPHARRQAIRESGQESNAGNQTIVRVFLLGOTPPED 189
DB 130 LLTDQPKCAKPFLLAIAKSLTPHARRQAIRESGQESNAGNQTIVRVFLLGOTPPED 189
QY 190 NHPDLSMLKFESEKHODILMNNYRDTFFNLSLKEVFLRWYVSTSCPTDFVFKGDDVDF 249
DB 190 NHPDLSMLKFESEKHODILMNNYRDTFFNLSLKEVFLRWYVSTSCPTDFVFKGDDVDF 249
QY 250 VNTTHILNLSLTKAKDLFIGDVHINAGPHRDKKLYIPEVYVSGLYPPYAGGGGF 309
DB 250 VNTTHILNLSLTKAKDLFIGDVHINAGPHRDKKLYIPEVYVSGLYPPYAGGGGF 309
QY 310 LYSGLHALRLYHITDOVHLYPIDDVYTGMLQKLGIVPEKHKGFRTFDEEKNKNICS 369
DB 310 LYSGLHALRLYHITDOVHLYPIDDVYTGMLQKLGIVPEKHKGFRTFDEEKNKNICS 369
QY 370 VDLMLVHSRKPQEMIDVWSQLOSAHLKC 397
DB 370 VDLMLVHSRKPQEMIDVWSQLOSAHLKC 397

RESULT 8

AA93536
ID AAB93536 standard; Protein: 373 AA.

XX

AC AAB93536:

XX

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:12898.

XX Human: primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 12898: 25377pp + CD ROM; English.
XX

CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC polynucleotide comprises at least 15 nucleotides and the combination
CC of the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 373 AA;

Query Match 94.5%; Score 2007; DB 22; Length 373;
Best Local Similarity 100.0%; Pred. No. 3.5e-206;
Matches 373; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 MEVSKSSQKNGKGEVLIIPKEFWKISTPPEAYWNREQEKLNQYNPILSMLTNGEA 84
DB 1 MEVSKSSQKNGKGEVLIIPKEFWKISTPPEAYWNREQEKLNQYNPILSMLTNGEA 60
QY 85 GRLSNLSHLNCEPDLRVTSVVTGFNNLPDRFKDFLLYLRCRNYSLIIDOPDKCAKPPFL 144
DB 61 GRLSNLSHLNCEPDLRVTSVVTGFNNLPDRFKDFLLYLRCRNYSLIIDOPDKCAKPPFL 120
QY 145 LLAIKSLTPHARRQAIRESGQESNAGNQTIVRVFLLGOTPPEDNHPDLSDMLKFESEK 204
DB 121 LLAIKSLTPHARRQAIRESGQESNAGNQTIVRVFLLGOTPPEDNHPDLSDMLKFESEK 180
QY 205 HQDILMNNYRDTFFNLSLKEVFLRWYVSTSCPTDFVFKGDDVDFVNTTHILNLSL 264
DB 181 HQDILMNNYRDTFFNLSLKEVFLRWYVSTSCPTDFVFKGDDVDFVNTTHILNLSL 240
QY 265 TKADLFIGDVHINAGPHRDKKLYIPEVYVSGLYPPYAGGGGFYSGLHALRLYHITD 324
DB 241 TKADLFIGDVHINAGPHRDKKLYIPEVYVSGLYPPYAGGGGFYSGLHALRLYHITD 300
QY 325 QVHLYPIDDVYTGMLQKLGIVPEKHKGFRTFDEEKNKNICSYYVDLMLVHSRKPQEMI 384
DB 301 QVHLYPIDDVYTGMLQKLGIVPEKHKGFRTFDEEKNKNICSYYVDLMLVHSRKPQEMI 360

```

QY 385 DIWSQLOSAHLKC 397
Db 361 DIWSQLOSAHLKC 373
RESULT 9
AAY84641
ID AAY84641 standard; Protein: 397 AA.
AC AAY84641;
XX
DT 25-JUL-2000 (first entry)
XX
DE Amino acid sequence of a murine LIG46 polypeptide.
XX
KW Leptin; LIG46; body weight; leptin inducible gene; obesity; cachexia;
KW LIG56; Tgtp; LRG-47; RC10-II; Stral3.
XX
OS Mus sp.
XX
PN W0200015926-A2.
XX
PD 23-MAR-2000.
XX
PF 10-SEP-1999; 99WO-US20722.
XX
PR 10-SEP-1998; 98US-0150857.
PR 29-OCT-1998; 98US-0106378.
PR 19-NOV-1998; 98US-0195896.
PR 15-APR-1999; 99US-0292228.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI White D, Zhou J, Tartaglia LA;
XX
DR WPI: 2000-271461/23.
DR N-PSDB; AAA12714.
XX
PT Method for determining compounds which modulate body weight and can be
PT used to treat e.g. obesity comprises measuring the activity of leptin
PT inducible genes .
XX
PS Claim 23; Fig 1A-B; 123pp; English.
XX
CC The present sequence represents a leptin induced LIG46 polypeptide. The
CC specification describes a method for determining whether a compound can
CC be used to modulate body weight by measuring the activity of leptin
CC inducible genes, such as LIG46. The method can be used to specifically
CC detect LIG46 and LIG56 nucleic acid molecules relative to other nucleic
CC acid molecules encoding galactosyltransferases or GMP-binding proteins.
CC They can also be used in diagnostic assays to identify the presence or
CC absence of a genetic lesion or mutation characterized by aberrant
CC modification or mis-regulation of the genes or aberrant post-
CC translational modification of the proteins. LIG46 and LIG56 proteins
CC and nucleic acid molecules can be used to treat obesity or cachexia.
CC LIG46 and LIG56 antagonists are used to treat obesity and LIG46 and
CC LIG56 agonists are used to treat low body weight. The leptin inducible
CC genes i.e. LIG46, LIG56, Tgtp, LRG-47, RC10-II and Stral3 can be used
CC to determine whether a compound modulates body weight and can then be
CC used to treat obesity or cachexia or low body weight.
XX
SQ Sequence 397 AA:
Query Match 88.28; Score 1872; DB 21; Length 397;
Best Local Similarity 87.48; Pred. No. 1.2e-191;
Matches 347; Conservative 19; Mismatches 31; Indels 0; Gaps 0;
QY 1 MSVGRRRIKLLGILMANVFIYFIMEVSKSSQKNGKEVILPKKFKWIKSTPPPAYWN 60
1 MSVGRRRKLLGILMANVFIYILVEVSKNSQDKNGKGVILPKKFKWPPSTPRAYWN 60
Db 1 MSVGRRRKLLGILMANVFIYILVEVSKNSQDKNGKGVILPKKFKWPPSTPRAYWN 60
QY 61 REQEKLNQYNPILSMILNTOTGEAGRLSNISHLNYPCEPDLRVTSVVTGFFNNLPDRFKDEL 120

```

CC fate specification, it is envisioned that wild type or mutant forms of
CC mammalian BRAINAC and/or EGHEAD can be used to alter epithelial cell
CC adhesion. This should be useful in treating many diseases which present
CC problems of cell viability, cell growth regulation and cell fate
CC specification. For example, these proteins, or active fragments or
CC analogues of these proteins and these genes can be used to treat
CC diseases such as cancer, psoriasis and other skin lesions, and nervous
CC system defects or diseases. The present sequence represents the murine
CC BRAINAC protein.
XX
XX
SQ Sequence 397 AA;

Query Match 88.2%; Score 1872; DB 21; Length 397;
Best Local Similarity 87.4%; Pred. No. 1.2e-191;
Matches 347; Conservative 19; Mismatches 31; Indels 0; Gaps 0;
Qy 1 MSVGRRRKILGLIMANVFYIFIMEYSSQKNGKGEVLPKEKFWKISTPPPEAYWN 60
Db 1 MSVGRRRKILGLIMANVFYIFIMEYSSQKNGKGEVLPKEKFWKISTPPPEAYWN 60
Qy 61 REQEKLRQYNPILSMLTNOTGEAGRLSNISHLNICEPDLRVTSVVTGFNNLPDRKDFL 120
Db 61 REQEKLRQYNPILSMLTNOTGEAGRLSNISHLNICEPDLRVTSVVTGFNNLPDRKDFL 120
Qy 121 LYLRCRNYSLIIDQPKCAKKPFLLLAIKSLTPHFARROAIRESWGSESNAGNQTWVRF 180
Db 121 LYLRCRNYSLIIDQPKCAKKPFLLLAIKSLTPHFARROAIRESWGSESNAGNQTWVRF 180
Qy 181 LGQTPPEDNHPDLSMDLKFESKHKQDILMWNRYDTFFNLSKEVLFRLWVSTCPDTEF 240
Db 181 LGQTPPEDNHPDLSMDLKFESKHKQDILMWNRYDTFFNLSKEVLFRLWVSTCPDTEF 240
Qy 241 VFKGDDVFNTHHILNLYNSLSKTRAKDLFGDVHINAGPHRDKKLYIPEVYVSGLY 300
Db 241 VFKGDDVFNTHHILNLYNSLSKTRAKDLFGDVHINAGPHRDKKLYIPEVYVSGLY 300
Qy 301 PPYAGGGFLYSGHLALRLYHITDQVHLYPIDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360
Db 301 PPYAGGGFLYSGHLALRLYHITDQVHLYPIDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360
Qy 361 KKNKNICSYVDLMVHSRKPQEMIDIWSQLQSAHLK 397
Db 361 KKNKNICSYVDLMVHSRKPQEMIDIWSQLQSAHLK 397

RESULT 11
AA: AAY69697
ID: AAY69697 standard; Protein: 397 AA.
AC: AAY69697;
XX 08-MAY-2000 (first entry)
XX Murine Brainiac protein.
XX Brainiac; murine; mammalian; expressed sequence tag; EST;
KW Drosophila melanogaster; Egghead; Notch; epithelial cell adhesion;
KW Fringe family; cell viability; growth regulation; cell fate; cancer;
KW psoriasis; skin lesion; nervous system disorder; developmental syndrome;
KW transgenic animal; drug screening.
XX OS Mus sp.
XX CA2255109-A1.
XX 17-JUN-1999.
XX 17-DEC-1998; 98CA-2255109.
XX 17-DEC-1997; 97CA-2225126.
XX (HSCR-) HSC RES & DEV LP.
XX

PI Egan SE;
XX WPI: 2000-148082/14.
DR N-PSDB; AAZ87185.
XX
XX New nucleic acids encoding a murine and human Brainiac protein, useful
PT for detecting somatic or germline DNA lesions which are responsible for
PT developmental syndromes or diseases including cancer.
XX
XX Claim 13; Page 28; 40pp; English.

CC This sequence represents murine Brainiac protein. The cDNA encoding this
CC protein was isolated from a mouse mammary gland cDNA library via the use
CC of a probe generated via PCR from a variety of murine tissues. The PCR
CC primers used for probe generation (AAZ87187-287188) were based on EST
CC (expressed sequence tag) sequences with significant similarity to
CC Drosophila melanogaster Brainiac. Drosophila Brainiac and Egghead
CC proteins regulate adhesion between epithelial cells, this activity being
CC dependent on the presence of Notch. Drosophila Brainiac is a secreted
CC protein, and has sequence similarities with the Drosophila Fringe
CC proteins. Brainiac/egghead-mediated epithelial cell adhesion is
CC required for cell viability, cell growth regulation and cell fate
CC specification. Wild-type or mutant forms of mammalian Brainiac
CC proteins may therefore be used to alter epithelial cell adhesion
CC in a mammal. Mammalian Brainiac proteins, active fragments, analogues,
CC and nucleic acids may be used to treat diseases such as cancer,
CC psoriasis and other skin lesions, and nervous system disorders.
CC Mammalian Brainiac nucleic acids may also be used to detect somatic
CC or germline DNA lesions which are responsible for developmental
CC syndromes or diseases including cancer. The mammalian Brainiac proteins
CC and fragments or its analogues are useful as antigens in immunoassays
CC including enzyme-linked immunosorbent assays (ELISA), radioimmunoassays
CC (RIA) and other non-enzyme linked antibody binding assays. Non-human
CC transgenic animals comprising nucleotide sequences encoding human
CC Brainiac protein (AAY69698) can be used as animal models for the study
CC of mammalian Brainiac gene function, for the screening of candidate
CC compounds and for the evaluation of potential therapeutic interventions.

XX Sequence 397 AA;

Query Match 88.2%; Score 1872; DB 21; Length 397;
Best Local Similarity 87.4%; Pred. No. 1.2e-191;
Matches 347; Conservative 19; Mismatches 31; Indels 0; Gaps 0;
Qy 1 MSVGRRRKILGLIMANVFYIFIMEYSSQKNGKGEVLPKEKFWKISTPPPEAYWN 60
Db 1 MSVGRRRKILGLIMANVFYIFIMEYSSQKNGKGEVLPKEKFWKISTPPPEAYWN 60
Qy 61 REQEKLRQYNPILSMLTNOTGEAGRLSNISHLNICEPDLRVTSVVTGFNNLPDRKDFL 120
Db 61 REQEKLRQYNPILSMLTNOTGEAGRLSNISHLNICEPDLRVTSVVTGFNNLPDRKDFL 120
Qy 121 LYLRCRNYSLIIDQPKCAKKPFLLLAIKSLTPHFARROAIRESWGSESNAGNQTWVRF 180
Db 121 LYLRCRNYSLIIDQPKCAKKPFLLLAIKSLTPHFARROAIRESWGSESNAGNQTWVRF 180
Qy 181 LGQTPPEDNHPDLSMDLKFESKHKQDILMWNRYDTFFNLSKEVLFRLWVSTCPDTEF 240
Db 181 LGQTPPEDNHPDLSMDLKFESKHKQDILMWNRYDTFFNLSKEVLFRLWVSTCPDTEF 240
Qy 241 VFKGDDVFNTHHILNLYNSLSKTRAKDLFGDVHINAGPHRDKKLYIPEVYVSGLY 300
Db 241 VFKGDDVFNTHHILNLYNSLSKTRAKDLFGDVHINAGPHRDKKLYIPEVYVSGLY 300
Qy 301 PPYAGGGFLYSGHLALRLYHITDQVHLYPIDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360
Db 301 PPYAGGGFLYSGHLALRLYHITDQVHLYPIDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360
Qy 361 KKNKNICSYVDLMVHSRKPQEMIDIWSQLQSAHLK 397
Db 361 KKNKNICSYVDLMVHSRKPQEMIDIWSQLQSAHLK 397

RESULT 12

AAG66115
 ID AAG66115 standard; Protein; 397 AA.
 AC AAG66115;
 XX 13-MAR-2002 (first entry)
 DT 13-MAR-2002 (first entry)
 XX Murine LIG46 polypeptide.
 DE
 XX Leptin; LIG46; LIG56; leptin; induced gene; Tgtp; LRG-47; RC10-II; Stral3;
 KW anorectic; anabolic; antiseize therapy; mouse.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..30
 FT /note= "signal peptide"
 FT Protein 31..397
 FT /note= "mature protein"
 XX
 PN US2001024808-A1.
 XX
 PD 27-SEP-2001.
 XX
 PF 12-MAR-2001; 2001US-0804357.
 XX
 PR 29-OCT-1998; 98US-106378P.
 PR 19-NOV-1998; 98US-0195896.
 PR 10-SEP-1998; 98US-0150857.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 XX White D, Zhou J, Tartaglia LA;
 PI WPI: 2001-624963/72.
 DR N-PSDB; AAI67865, AAI67866.
 XX
 XX An isolated polypeptide useful for treating a weight disorder and for
 PT screening compounds that may inhibit protein activity comprises a
 PT leptin-induced protein.
 XX
 PS Claim 22; Fig 1: 46pp; English.
 XX
 CC The invention relates to genes whose expression are induced by leptin.
 CC LIG46 and LIG56 are novel leptin induced genes (LIG), while four other
 CC genes such as Tgtp, LRG-47, RC10-II and Stral3 have been previously
 CC identified. The leptin induced proteins can be expressed by standard
 CC recombinant methodology. The proteins and encoding polynucleotides may
 CC be used in screening assays to identify compounds that may bind to it.
 CC Administering a molecule, e.g., an antiseize molecule, which reduces
 CC expression of activity of protein selected from LIG46, LIG56, Tgtp,
 CC LRG-47, RC10-II, and Stral3 may be used to treat a weight disorder. The
 CC method may further comprise administering leptin. The present sequence
 CC represents a murine LIG46 protein.
 XX
 SQ Sequence 397 AA;
 Query Match 88.2%; Score 1872; DB 22; Length 397;
 Best Local Similarity 87.4%; Pred. No. 1.2e-191;
 Matches 347; Conservative 19; Mismatches 31; Indels 0; Gaps 0;
 QY 1 MSVGRRIKLLGILMANVFIYIMEVSKSSQKNGKGVFIIPKFKFKIPTTPPAYN 60
 DB 1 MSVGRRRVKKLLGILMANVFIYIVEVSKNSQDKNGKGVFIIPKFKFKIPTTPPAYN 60
 QY 61 REQELKRNXPILSMITNCTGAGRLSNISHLNYCEPDLRTVTSVVTGNNLPDRKDEL 120
 DB 61 REQELKRNWNPILNRVANOTGELATSPNTSHLSYCEPDSVTMTAVTDFNLPDRKDEL 120
 QY 121 LYLCRNYSLLIDQPKCAKKPFLLLAKSLTPHFARRQAIRESWGQESNAGNQTIVRVF 180
 DB 121 LYLCRNYSLLIDQPKCAKKPFLLLAKSLTPHFARRQAIRESWGQETNNGNQTIVRVF 180

QY 181 ILGQTPEDNHPDLSOMLKFSEKSHODILMNWYRDTFFNLKSLKEVFLRWVSTSCPDTEF 240
 DB 181 LLCKTPEDNHPDLSOMLKFESDKHODILMNWYRDTFFNLKSLKEVFLRWVSTSCPDTEF 240
 QY 241 VFKGDDVFNTHHILNLSKTKAKOLFICGVTHNAGPHRDKKLKYYIPEVYVSGLY 300
 DB 241 VFKGDDVFNTHHILNLSKTKAKOLFICGVTHNAGPHRDKKLKYYIPEVYVSGLY 300
 QY 301 PPVAGGGFLYSGHLALRLVHTDQVHLPIPDVYTGMCLOKLGVLPEKHKGFRFTDIEE 360
 DB 301 PPVAGGGFLYSGHLALRLVHTDQVHLPIPDVYTGMCLOKLGVLPEKHKGFRFTDIEE 360
 QY 361 KKNKNICSYVDLMLVHSRKPQEMIDIWSQLSAHLKC 397
 DB 361 KKNKNICSYIDLMLVHSRKPQEMIDIWSQLSPNLKC 397
 RESULT 13
 AAE29092
 ID AAE29092 standard; Protein; 397 AA.
 XX AAE29092;
 XX 27-JAN-2003 (first entry)
 DT 27-JAN-2003 (first entry)
 DE Murine LIG46 protein.
 XX
 KW LIG46; leptin; low body weight; chromosomal mapping; tissue typing;
 KW forensic biology; transgenic; gene therapy; antianorectic; murine.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..32
 FT /label= Signal_peptide
 FT Protein 33..397
 FT /note= "Murine mature LIG46 protein"
 FT Modified-site 3..6
 FT /note= "Amidation site"
 FT Modified-site 30..33
 FT /note= "N-glycosylation site"
 FT Modified-site 31..34
 FT /note= "Casein kinase II phosphorylation site"
 FT Domain 33..302
 FT /note= "Extracellular domain"
 FT Modified-site 54..56
 FT /note= "Protein kinase C phosphorylation site"
 FT Modified-site 79..82
 FT /note= "N-glycosylation site"
 FT Modified-site 89..92
 FT /note= "N-glycosylation site"
 FT Modified-site 94..97
 FT /note= "Casein kinase II phosphorylation site"
 FT Modified-site 115..122
 FT /note= "Tyrosine kinase phosphorylation site"
 FT Modified-site 127..173
 FT /note= "N-glycosylation site"
 FT Modified-site 185..188
 FT /note= "Casein kinase II phosphorylation site"
 FT Modified-site 202..204
 FT /note= "Protein kinase C phosphorylation site"
 FT Modified-site 219..222
 FT /note= "N-glycosylation site"
 FT Modified-site 221..224
 FT /note= "Casein kinase II phosphorylation site"
 FT Modified-site 221..223
 FT /note= "Protein kinase C phosphorylation site"
 FT Modified-site 234..237
 FT /note= "Casein kinase II phosphorylation site"
 FT Domain 303..320
 FT /note= "Transmembrane domain"
 FT Domain 321..397

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 22:49:49 ; Search time 38.5 Seconds
(without alignments)
1636.739 Million cell updates/sec

Title: US-09-804-357B-2

Perfect score: 2129

Sequence: 1 MSVRRRVKLLGILMANVF.....RKPQEMDIWSQLQSPNLKC 397

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_l9Jun03:*

```
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2119	99.5	397	21	AAV84641
2	2119	99.5	397	21	AAV79953
3	2119	99.5	397	21	AAV69697
4	2119	99.5	397	22	AAV66115
5	2119	99.5	397	23	AAE29092
6	2113	99.2	397	24	ABB82648
7	1977	92.9	367	22	AAV66116
8	1862	87.5	397	19	AAW80212
9	1862	87.5	397	21	AAV84683

10	1862	87.5	397	22	AAV66118	Human LIG46 polype
11	1862	87.5	397	23	AAE29093	Human LIG46 protei
12	1862	87.5	397	24	ABB82649	Human beta3Gnt po
13	1844	86.6	413	22	AAV49748	Human beta 1,3-N-a
14	1784	83.8	397	21	AAV69698	Human Brainiac pro
15	1753	82.3	373	22	AAV93536	Human Brainiac pro
16	1680	78.9	397	21	AAV79954	Human protein sequ
17	980	46.0	197	23	AAE15936	Human brainiac pro
18	656	30.8	352	20	AAV06462	Human Brainiac-3.
19	656	30.8	352	24	ABR47731	Human secreted pro
20	656	30.8	353	20	AAV36224	Human secreted pro
21	654.5	30.7	401	23	AAE22148	Human TNFR-10 pro
22	647.5	30.4	378	21	AAV40333	Human PRO4344 prot
23	647.5	30.4	378	22	AAV49751	Human beta 1,3-N-a
24	647.5	30.4	378	23	ABG34047	Human Pro peptide
25	646	30.3	384	23	ABG32374	The beta-1,3-galac
26	635.5	29.8	378	21	AAV07435	Novel human enzyme
27	635.5	29.8	378	23	ABG32941	Human galactosyltr
28	635.5	29.8	378	23	AAE24688	Human zissp6 prote
29	633.5	29.8	334	22	ABG20251	Novel human diagno
30	631.5	29.7	364	23	ABU65191	Human NOV106a prot
31	623	29.3	415	23	ABG30984	Human galactosyltr
32	622.5	29.2	372	22	AAV49750	Human beta 1,3-N-a
33	619.5	29.1	372	19	AAV64558	Human epidermoid c
34	619.5	29.1	372	22	AAU29167	Human PRO polypept
35	618.5	29.1	372	22	AAV49749	Human beta 1,3-N-a
36	619.5	29.1	372	23	ABB09716	Amino acid sequenc
37	619.5	29.1	372	23	AAU11272	Human beta1,3-N-ac
38	619.5	29.1	372	24	ABU71255	Human PRO1266 prot
39	619.5	29.1	372	24	ABU65712	Human secreted/tra
40	619.5	29.1	372	24	ABU66045	Novel human secret
41	619.5	29.1	372	24	ABU67549	Human secreted/tra
42	619.5	29.1	372	24	ABR01801	Human cancer-relat
43	619.5	29.1	372	24	ABU65407	Human PRO polypept
44	619.5	29.1	372	24	ABU58543	Human PRO polypept
45	619.5	29.1	372	24	ABU56079	Human secreted/tra

ALIGNMENTS

RESULT 1

AAV84641

ID AAV84641 standard; Protein; 397 AA.

AC AAV84641;

DT 25-JUL-2000 (first entry)

DE Amino acid sequence of a murine LIG46 polypeptide.

KW Leptin; LIG46; body weight; leptin inducible gene; obesity; cachexia;
LIG56; Tgtp; LRG-47; RC10-II; Stral3.

OS Mus sp.

PN WO200015826-A2.

PP 23-MAR-2000.

PF 10-SEP-1999; 99WO-US20722.

PR 10-SEP-1998; 98US-0150857.

PR 29-OCT-1998; 98US-0106378.

PR 19-NOV-1998; 98US-0195896.

PR 15-APR-1999; 99US-0292228.

XX (MILL-) MILLENNIUM PHARM INC.

XX White D, Zhou J, Tartaglia LA;

XX WPI; 2000-271461/23.

DR N-PSDB; AAA12714.

XX Method for determining compounds which modulate body weight and can be
 PT used to treat e.g. obesity comprises measuring the activity of leptin
 PT inducible genes

XX Claim 23: Fig 1A-B; 123pp; English.

XX The present sequence represents a leptin induced LIG46 polypeptide. The
 CC specification describes a method for determining whether a compound can
 CC be used to modulate body weight by measuring the activity of leptin
 CC inducible genes, such as LIG46. The method can be used to specifically
 CC detect LIG46 and LIG56 nucleic acid molecules relative to other nucleic
 CC acid molecules encoding galactosyltransferases or GTP-binding proteins.
 CC They can also be used in diagnostic assays to identify the presence or
 CC absence of a genetic lesion or mutation characterized by aberrant
 CC modification or mis-regulation of the genes or aberrant post-
 CC translational modification of the proteins. LIG46 and LIG56 proteins
 CC and nucleic acid molecules can be used to treat obesity or cachexia.
 CC LIG46 and LIG56 antagonists are used to treat obesity and LIG46 and
 CC LIG56 agonists are used to treat low body weight. The leptin inducible
 CC genes i.e. LIG46, LIG56, Tgtp, LRG-47, RC10-11 and Stral3 can be used
 CC to determine whether a compound modulates body weight and can then be
 CC used to treat obesity or cachexia or low body weight.

XX Sequence 397 AA:

Query Match 99.58; Score 2119; DB 21; Length 397;
 Best Local Similarity 99.58; Pred. No. 1e-223;
 Matches 395; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSVGRRRVKLLGILMANVFIYLIIVEVSKNSODKNGKGGVITPKKFKWKPSTPRAYWN 60
 DB 1 MSVGRRRVKLLGILMANVFIYLIIVEVSKNSODKNGKGGVITPKKFKWKPSTPRAYWN 60

QY 61 REQEKLNRWNPILNRVANOTGELATSPNTSHLSYCEPOSTVMTAVTDNNLPDRKDFL 120
 DB 61 REQEKLNRWNPILNRVANOTGELATSPNTSHLSYCEPOSTVMTAVTDNNLPDRKDFL 120

QY 121 LYLRCRNYSLIDOPKKCAKPPFLLAISKSLIPHARRQAIRESWGRETNGVQTVVYRVF 180
 DB 121 LYLRCRNYSLIDOPKKCAKPPFLLAISKSLIPHARRQAIRESWGRETNGVQTVVYRVF 180

QY 181 LLGKTPPEDNHPDLSMDLKFSKAKDLFIMGDVHINAGPHRDKKLYIPEVFTGYV 300
 DB 181 LLGKTPPEDNHPDLSMDLKFSKAKDLFIMGDVHINAGPHRDKKLYIPEVFTGYV 300

QY 301 PPYAGGGFLYSGPALRLYSATSRVHLXPIDDVYTCMCLQKLGVLPEKHKGFTFIDIE 360
 DB 301 PPYAGGGFLYSGPALRLYSATSRVHLXPIDDVYTCMCLQKLGVLPEKHKGFTFIDIE 360

QY 361 KKKKNICSYIDLMLVHSRKPQEMIDWSQLQSNLKC 397
 DB 361 KKKKNICSYIDLMLVHSRKPQEMIDWSQLQSNLKC 397

RESULT 2
 AAY79953
 ID AAY79953 standard; Protein; 397 AA.

XX AAY79953;
 AC AAY79953;
 XX 12-MAY-2000 (first entry)
 DT 12-MAY-2000 (first entry)
 XX Marine brainiac protein.

XX Brainiac; egghead; cell adhesion; cytostatic; dermatological;
 KW neuroactive; cell viability; cell growth regulation; psoriasis; cancer;
 KW cell fate specification; skin lesion; nervous system defect.

OS Mus sp.
 XX CA2225126-A1.
 XX 17-JUN-1999.
 XX 17-DEC-1997; 97CA-2225126.
 XX 17-DEC-1997; 97CA-2225126.
 XX (HSCR-) KSC RES & DEV LP.
 PI Egan SE;
 XX WPI: 2000-161481/15.
 DR N-PSDB; AAZ88486.
 XX Mammalian EGGHEAD and BRAINIAC proteins which mediate cell to cell
 PT adhesion and may be used to treat cancer, psoriasis and other skin
 PT lesions and nervous system defects or diseases -
 XX Example 1: Page 25; 30pp; English.

XX The present invention describes mammalian EGGHEAD and BRAINIAC proteins,
 CC which mediate cell-to-cell adhesion. The mammalian proteins are similar
 CC to the EGGHEAD and BRAINIAC genes of Drosophila melanogaster. BRAINIAC
 CC and EGGHEAD genes regulate adhesion between epithelial cells and require
 CC the presence of the Notch protein for this activity. Mutational
 CC disruption of BRAINIAC, EGGHEAD or Notch results in the loss of follicle
 CC epithelial cell adhesion, thereby reducing the efficiency of signalling
 CC factor receptor. The BRAINIAC and EGGHEAD proteins have cytostatic,
 CC dermatological and neuroactive properties. As the BRAINIAC and EGGHEAD
 CC proteins are involved in adhesion between epithelial cells, and as this
 CC adhesion is required for cell viability, cell growth regulation and cell
 CC fate specification, it is envisioned that wild type or mutant forms of
 CC mammalian BRAINIAC and/or EGGHEAD can be used to alter epithelial cell
 CC adhesion. This should be useful in treating many diseases which present
 CC problems of cell viability, cell growth regulation and cell fate
 CC specification. For example, these proteins, or active fragments or
 CC analogues of these proteins and these genes can be used to treat
 CC diseases such as cancer, psoriasis and other skin lesions, and nervous
 CC system defects or diseases. The present sequence represents the murine
 CC BRAINIAC protein.

XX Sequence 397 AA:

Query Match 99.58; Score 2119; DB 21; Length 397;
 Best Local Similarity 99.58; Pred. No. 1e-223;
 Matches 395; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSVGRRRVKLLGILMANVFIYLIIVEVSKNSODKNGKGGVITPKKFKWKPSTPRAYWN 60
 DB 1 MSVGRRRVKLLGILMANVFIYLIIVEVSKNSODKNGKGGVITPKKFKWKPSTPRAYWN 60

QY 61 REQEKLNRWNPILNRVANOTGELATSPNTSHLSYCEPOSTVMTAVTDNNLPDRKDFL 120
 DB 61 REQEKLNRWNPILNRVANOTGELATSPNTSHLSYCEPOSTVMTAVTDNNLPDRKDFL 120

QY 121 LYLRCRNYSLIDOPKKCAKPPFLLAISKSLIPHARRQAIRESWGRETNGVQTVVYRVF 180
 DB 121 LYLRCRNYSLIDOPKKCAKPPFLLAISKSLIPHARRQAIRESWGRETNGVQTVVYRVF 180

QY 181 LLGKTPPEDNHPDLSMDLKFSKAKDLFIMGDVHINAGPHRDKKLYIPEVFTGYV 300
 DB 181 LLGKTPPEDNHPDLSMDLKFSKAKDLFIMGDVHINAGPHRDKKLYIPEVFTGYV 300

QY 301 PPYAGGGFLYSGPALRLYSATSRVHLXPIDDVYTCMCLQKLGVLPEKHKGFTFIDIE 360
 DB 301 PPYAGGGFLYSGPALRLYSATSRVHLXPIDDVYTCMCLQKLGVLPEKHKGFTFIDIE 360

Db 301 PPYAGGGFLYSGPLALRLYSATSRVHLYPIDDVYTGMCLOKGLGLVPEKHGKGFRTFIDE 360

Qy 361 KNKNKICSYIDLMLVHSRKPQEMIDIWSQLSPNLKC 397
|||||
Db 361 KNKNKICSYIDLMLVHSRKPQEMIDIWSQLSPNLKC 397

RESULT 3
AAV69697
ID AAV69697 standard; Protein: 397 AA.
AC AAV69697;
XX
XX 08-MAY-2000 (first entry)
DT
DE Murine Brainiac protein.
XX
XX Brainiac; murine; mammalian; expressed sequence tag; EST;
KW Drosophila melanogaster; Egghead; Notch; epithelial cell adhesion;
KW Fringe family; cell viability; growth regulation; cell fate; cancer;
KW Psoriasis; skin lesion; nervous system disorder; developmental syndrome;
KW transgenic animal; drug screening.
XX
OS Mus sp.
XX
XX CA2255109-A1.
PN
XX
XX 17-JUN-1999.
PD
XX
XX 17-DEC-1998; 98CA-2255109.
PF
XX
XX 17-DEC-1997; 97CA-2225126.
PR
XX
XX (HSCR-) HSC RES & DEV LP.
PA
XX
XX Egan SE;
PI
XX
XX WPI; 2000-148082/14.
DR
XX
XX N-PSDB; AAZ87185.
XX
XX New nucleic acids encoding a murine and human Brainiac protein, useful
PT for detecting somatic or germline DNA lesions which are responsible for
PT developmental syndromes or diseases including cancer -
XX
XX Claim 13; Page 28; 40pp; English.
XX
XX This sequence represents murine Brainiac protein. The cDNA encoding this
CC protein was isolated from a mouse mammary gland cDNA library via the use
CC of a probe generated via PCR from a variety of murine tissues. The PCR
CC primers used for probe generation (AA287187-287188) were based on EST
CC (expressed sequence tag) sequences with significant similarity to
CC Drosophila melanogaster Brainiac. Drosophila Brainiac and Egghead
CC proteins regulate adhesion between epithelial cells, this activity being
CC dependent on the presence of Notch. Drosophila Brainiac is a secreted
CC protein, and has sequence similarities with the Drosophila Fringe
CC proteins. Brainiac/Egghead-mediated epithelial cell adhesion is
CC required for cell viability, cell growth regulation and cell fate
CC specification. Wild-type or mutant forms of mammalian Brainiac
CC proteins may therefore be used to alter epithelial cell adhesion
CC in a mammal. Mammalian Brainiac proteins, active fragments, analogues,
CC and nucleic acids may be used to treat diseases such as cancer,
CC psoriasis and other skin lesions, and nervous system disorders.
CC Mammalian Brainiac nucleic acids may also be used to detect somatic
CC or germline DNA lesions which are responsible for developmental
CC syndromes or diseases including cancer. The mammalian Brainiac proteins
CC and fragments or its analogues are useful as antigens in immunoassays
CC including enzyme-linked immunosorbent assays (ELISA), radioimmunoassays
CC (RIA) and other non-enzyme linked antibody binding assays. Non-human
CC transgenic animals comprising nucleotide sequences encoding human
CC Brainiac protein (AAV69698) can be used as animal models for the study
CC of mammalian Brainiac gene function, for the screening of candidate
CC compounds and for the evaluation of potential therapeutic interventions.
XX

SQ Sequence 397 AA:
Query Match 99.5%; Score 2119; DB 21; Length 397;
Best Local Similarity 99.5%; Pred. No. 1e-223; Mismatches 0; Gaps 0;
Matches 395; Conservative 0; Indels 0; Gaps 0;

Qy 1 MSVGRRRVKLLGILMMANVFYLIIVEVSKNSQDKNGKGGVLIIPKEKFWKPPSTPRAYN 60
|||||
Db 1 MSVGRRRVKLLGILMMANVFYLIIVEVSKNSQDKNGKGGVLIIPKEKFWKPPSTPRAYN 60
|||||

Qy 61 REQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNLPDRFKDFL 120
|||||
Db 61 REQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNLPDRFKDFL 120
|||||

Qy 121 LYLRGRNYSLLIDQPKCAKPFLLAIAKSLIPHARRQAIRRESWGRETNYGNQTVVRVF 180
|||||
Db 121 LYLRGRNYSLLIDQPKCAKPFLLAIAKSLIPHARRQAIRRESWGRETNYGNQTVVRVF 180
|||||

Qy 181 LLGKTPPENHDPLSDMLKFESDKHODILMMNYRTFFNLSLKEVFLRWVSTSCPDADF 240
|||||
Db 181 LLGKTPPENHDPLSDMLKFESDKHODILMMNYRTFFNLSLKEVFLRWVSTSCPDADF 240
|||||

Qy 241 VFKGDDVFNTHHILNYLSLSKSKAKDLFTGDVHNAGPHRDKKLYIPEVFTGVY 300
|||||
Db 241 VFKGDDVFNTHHILNYLSLSKSKAKDLFTGDVHNAGPHRDKKLYIPEVFTGVY 300
|||||

Qy 301 PPYAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKGLGLVPEKHGKGFRTFIDE 360
|||||
Db 301 PPYAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKGLGLVPEKHGKGFRTFIDE 360
|||||

Qy 361 KNKNKICSYIDLMLVHSRKPQEMIDIWSQLSPNLKC 397
|||||
Db 361 KNKNKICSYIDLMLVHSRKPQEMIDIWSQLSPNLKC 397
|||||

RESULT 4
AAG66115
ID AAG66115 standard; Protein: 397 AA.
XX
AC AAG66115;
XX
DT 13-MAR-2002 (first entry)
XX
XX Marine LIG46 polypeptide.
DE
XX
KW Leptin; LIG46; leptin induced gene; Tgtp; LRG-47; RC10-II; Stral3;
KW anorectic; anabolic; antisense therapy; mouse.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT Peptide 1..30 /note= "signal peptide"
FT Protein 31..397 /note= "mature protein"
XX
XX US2001024808-A1.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 12-MAR-2001; 2001US-0804357.
XX
XX 29-OCT-1998; 98US-106378P.
PR 19-NOV-1998; 98US-0195896.
PR 10-SEP-1998; 98US-0150857.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX White D, Zhou J, Tartaglia LA;
XX WPI; 2001-624963/72.
DR N-PSDB; AA167865, AA167866.
XX

PT An isolated polypeptide useful for treating a weight disorder and for
 PT screening compounds that may inhibit protein activity comprises a
 PT leptin-induced protein

XX Claim 22: Fig 1: 46pp; English.

XX The invention relates to genes whose expression are induced by leptin.
 CC LIG46 and LIG56 are novel leptin induced genes (LIG), while four other
 CC genes such as Tgtp, LRG-47, RC10-II and Stral3 have been previously
 CC identified. The leptin induced proteins can be expressed by standard
 CC recombinant methodology. The proteins and encoding polynucleotides may
 CC be used in screening assays to identify compounds that may bind to it.
 CC Administering a molecule, e.g., an antisense molecule, which reduces
 CC expression of activity of protein selected from LIG46, LIG56, Tgtp,
 CC LRP-47, RC10-II, and Stral3 may be used to treat a weight disorder. The
 CC method may further comprise administering leptin. The present sequence
 CC represents a murine LIG46 protein.

XX Sequence 397 AA;

Query Match 99.5%; Score 2119; DB 22; Length 397;

Best Local Similarity 99.5%; Pred. No. 1e-223;

Matches 395; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSVGRRVKLLGLTMANVFYILVEVSKNSQDKNGGVIIPKEKFKPSPRAYWN 60
 DB 1 MSVGRRVKLLGLTMANVFYILVEVSKNSQDKNGGVIIPKEKFKPSPRAYWN 60
 QY 61 REEKLNRWYNPILNRVANOTGELATSPNTHLSYCEPDSTVMTAVTFNPLDRFKDFL 120
 DB 61 REEKLNRWYNPILNRVANOTGELATSPNTHLSYCEPDSTVMTAVTFNPLDRFKDFL 120
 QY 121 LYLRCRNYSLIDQPKCAKPFLLAIKSLIPHARRQAIFRESWGRTNYSNOTVVRVF 180
 DB 121 LYLRCRNYSLIDQPKCAKPFLLAIKSLIPHARRQAIFRESWGRTNYSNOTVVRVF 180
 QY 181 LKATPEDNHPDLSMLKFESDKHODILMWNRYRTFFNLSLKEYFLFLRWYSTSCPAEF 240
 DB 181 LKATPEDNHPDLSMLKFESDKHODILMWNRYRTFFNLSLKEYFLFLRWYSTSCPAEF 240
 QY 241 VFKGDDDFVFNTHILNLSLSKSKAKDLFIGDVHINAGPHRDKLKYIPEVFYGVY 300
 DB 241 VFKGDDDFVFNTHILNLSLSKSKAKDLFIGDVHINAGPHRDKLKYIPEVFYGVY 300
 QY 301 PPVAGGGFLYSGFALLRLYSATSRVHLYPIDVYTGMCLOKGLVPEKHKGFRTFDEE 360
 DB 301 PPVAGGGFLYSGFALLRLYSATSRVHLYPIDVYTGMCLOKGLVPEKHKGFRTFDEE 360
 QY 361 KNKKNICSYIDLMLVHSRKPQEMID:WSOLQSPNLKC 397
 DB 361 KNKKNICSYIDLMLVHSRKPQEMID:WSOLQSPNLKC 397

RESULT 5

AAE29092

XX ID AAE29092 standard; Protein: 397 AA.

XX AC AAE29092;

XX 27-JAN-2003 (first entry)

XX Murine LIG46 protein.

XX LIG46; leptin; low body weight; chromosomal mapping; tissue typing;

XX forensic biology; transgenic; gene therapy; antianorectic; murine.

XX Mus sp.

XX Key Location/Qualifiers

XX 1..32 /label= Signal_peptide

XX 33..397 /note= "Murine mature LIG46 protein"

FT Modified-site 3..6 /note= "Amidation site"
 FT Modified-site 30..33 /note= "N-glycosylation site"
 FT Modified-site 31..34 /note= "Casein kinase II phosphorylation site"
 FT Domain 33..302 /note= "Extracellular domain"
 FT Modified-site 54..56 /note= "Protein kinase C phosphorylation site"
 FT Modified-site 79..82 /note= "N-glycosylation site"
 FT Modified-site 89..92 /note= "N-glycosylation site"
 FT Modified-site 94..97 /note= "Casein kinase II phosphorylation site"
 FT Modified-site 115..122 /note= "Tyrosine kinase phosphorylation site"
 FT Modified-site 127..173 /note= "N-glycosylation site"
 FT Modified-site 185..188 /note= "Casein kinase II phosphorylation site"
 FT Modified-site 202..204 /note= "Protein kinase C phosphorylation site"
 FT Modified-site 219..222 /note= "N-glycosylation site"
 FT Modified-site 221..224 /note= "Casein kinase II phosphorylation site"
 FT Modified-site 221..223 /note= "Protein kinase C phosphorylation site"
 FT Modified-site 234..237 /note= "Casein kinase II phosphorylation site"
 FT Domain 303..320 /note= "Transmembrane domain"
 FT Domain 321..397 /note= "Cytoplasmic domain"
 FT Modified-site 323..325 /note= "Protein kinase C phosphorylation site"
 FT Modified-site 368..371 /note= "Casein kinase II phosphorylation site"
 FT Modified-site 377..379 /note= "Protein kinase C phosphorylation site"
 PN W0200274905-A2.
 XX 26-SEP-2002.
 XX 20-NOV-2001; 2001WO-US43345.
 XX 21-NOV-2000; 2000US-0717778.
 XX (MILLI-) MILLENNIUM PHARM INC.
 XX White DW, Zhou J, Tartaglia LA, Stricker-krongrad A, Clausen H;
 DR WPI: 2002-759886/82.
 DR N-PSDB; AAD46672.
 XX New isolated nucleic acid molecules, designated as leptin-induced genes
 PT 46 (LIG46), useful for treating a subject having a disorder
 PT characterized by undesirable level of LIG46 expression or activity,
 PT such as low body weight
 XX Claim 22; Fig 1; 90pp; English.
 XX The invention relates to LIG46, a gene whose expression is induced by
 CC leptin. LIG46 DNA and protein are useful in treating a subject having
 CC a disorder characterised by undesirable level of LIG46 expression or
 CC activity, such as low body weight. They are also useful in a screening
 CC assay, chromosomal mapping, tissue typing and forensic biology. The
 CC probes based on the LIG46 nucleotide sequence are useful for detecting
 CC transcripts or genomic sequences encoding the same or related proteins.
 CC The LIG46 polypeptides are useful as immunogens for raising anti-LIG46

CC antibodies. The host cells are useful for producing non-human transgenic
 CC animals. LIG46 DNA is used in gene therapy. The present sequence is
 CC murine LIG46 protein.
 XX
 SQ Sequence 397 AA;
 Query Match 99.5%; Score 2119; DB 23; Length 397;
 Best Local Similarity 99.5%; Pred. No. 1e-223;
 Matches 395; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MSVGRRRVKLLGILMANVFIYLIIVEVSKNSQDKNGKGGVLIIPKEKFWKPPSTPRAYWN 60
 DB 1 MSVGRRRVKLLGILMANVFIYLIIVEVSKNSQDKNGKGGVLIIPKEKFWKPPSTPRAYWN 60
 QY 61 REEKLNRWNPILNRVANQTGELATSPNTHLSYCEPDSTVMTAVTFNNLPDRFKDFL 120
 DB 61 REEKLNRWNPILNRVANQTGELATSPNTHLSYCEPDSTVMTAVTFNNLPDRFKDFL 120
 QY 121 LYLRCRNYSLIDOPKKCAKPFLLAIAKSLIPHARRQAIRESWGRETNGVQTVRVF 180
 DB 121 LYLRCRNYSLIDOPKKCAKPFLLAIAKSLIPHARRQAIRESWGRETNGVQTVRVF 180
 QY 181 LLGKTPPEDNHPDLSMDLKFESEKHODILMWNRYDTFFNLSEKVLFLRWVSTSCPDADF 240
 DB 181 LLGKTPPEDNHPDLSMDLKFESEKHODILMWNRYDTFFNLSEKVLFLRWVSTSCPDADF 240
 QY 241 VFKGDDVFNTHILNLSKSKAKOLFIDGVIHNAGPHRDKKLKYYIPEVFTGYV 300
 DB 241 VFKGDDVFNTHILNLSKSKAKOLFIDGVIHNAGPHRDKKLKYYIPEVFTGYV 300
 QY 301 PPVAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKGLVPEKHKGFTFDIEE 360
 DB 301 PPVAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKGLVPEKHKGFTFDIEE 360
 QY 361 KNKKNICSYIDLMLVHSRKPQEMIDIWSQLSPNLKC 397
 DB 361 KNKKNICSYIDLMLVHSRKPQEMIDIWSQLSPNLKC 397
 RESULT 6
 ABB82648
 ID ABB82648 standard; Protein: 397 AA.
 XX
 AC ABB82648;
 XX
 DT 19-FEB-2003 (first entry)
 DE Murine beta3Gnt polypeptide sequence.
 DE Beta3Gnt; transgenic; beta-1,3-N-acetylglucosaminyltransferase;
 KW disease system; mouse; enzyme.
 XX
 OS Mus sp.
 XX
 WO20027943-A2.
 XX
 PD 10-OCT-2002.
 XX
 PF 29-MAR-2002; 2002WO-US09645.
 XX
 PR 29-MAR-2001; 2001US-280706P.
 PR 28-MAR-2002; 2002US-0280706.
 XX
 XX (DELT-) DELTAGEN INC.
 PA Leventen MW, Phillips R;
 PI WPI; 2003-067437/06.
 XX N-PSDB; ABV75081.
 DR
 DR
 XX New transgenic mouse comprising a disruption in a beta3Gnt gene, as in
 PT vivo model to study various disease states or conditions in which
 PT beta3Gnt may be implicated or involved, such as abnormal cell growth,

PT cancer and metastasis -
 XX Disclosure; Fig 2; 55pp; English.
 XX
 CC The invention relates to a transgenic mouse comprising a disruption in a
 CC beta3Gnt (beta-1,3-N-acetylglucosaminyltransferase) gene, where there is
 CC no native expression of beta3Gnt gene. The transgenic mice may be used
 CC as in vivo model to study various disease states or conditions in which
 CC beta3Gnt may be implicated or involved, such as abnormal cell growth,
 CC cancer and metastasis, and to evaluate various treatments or to identify
 CC agents for treating disease states or conditions, such as anxiety or
 CC depression. Animal-based disease systems may be used to identify
 CC compounds capable of ameliorating disease symptoms, as test substrates
 CC for the identification of drugs, pharmaceuticals, therapies and
 CC interventions that may be effective in treating a disease or other
 CC phenotypic characteristic of the animal. The present sequence represents
 CC a mouse beta3Gnt polypeptide sequence.
 XX
 SQ Sequence 397 AA;
 Query Match 99.2%; Score 2113; DB 24; Length 397;
 Best Local Similarity 99.2%; Pred. No. 4.8e-223;
 Matches 394; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSVGRRRVKLLGILMANVFIYLIIVEVSKNSQDKNGKGGVLIIPKEKFWKPPSTPRAYWN 60
 DB 1 MSVGRRRVKLLGILMANVFIYLIIVEVSKNSQDKNGKGGVLIIPKEKFWKPPSTPRAYWN 60
 QY 61 REEKLNRWNPILNRVANQTGELATSPNTHLSYCEPDSTVMTAVTFNNLPDRFKDFL 120
 DB 61 REEKLNRWNPILNRVANQTGELATSPNTHLSYCEPDSTVMTAVTFNNLPDRFKDFL 120
 QY 121 LYLRCRNYSLIDOPKKCAKPFLLAIAKSLIPHARRQAIRESWGRETNGVQTVRVF 180
 DB 121 LYLRCRNYSLIDOPKKCAKPFLLAIAKSLIPHARRQAIRESWGRETNGVQTVRVF 180
 QY 181 LLGKTPPEDNHPDLSMDLKFESEKHODILMWNRYDTFFNLSEKVLFLRWVSTSCPDADF 240
 DB 181 LLGKTPPEDNHPDLSMDLKFESEKHODILMWNRYDTFFNLSEKVLFLRWVSTSCPDADF 240
 QY 241 VFKGDDVFNTHILNLSKSKAKOLFIDGVIHNAGPHRDKKLKYYIPEVFTGYV 300
 DB 241 VFKGDDVFNTHILNLSKSKAKOLFIDGVIHNAGPHRDKKLKYYIPEVFTGYV 300
 QY 301 PPVAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKGLVPEKHKGFTFDIEE 360
 DB 301 PPVAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKGLVPEKHKGFTFDIEE 360
 QY 361 KNKKNICSYIDLMLVHSRKPQEMIDIWSQLSPNLKC 397
 DB 361 KNKKNICSYIDLMLVHSRKPQEMIDIWSQLSPNLKC 397
 RESULT 7
 AAG66116
 ID AAG66116 standard; Protein: 367 AA.
 XX
 AC AAG66116;
 XX
 DT 13-MAR-2002 (first entry)
 DE Murine LIG46 mature polypeptide sequence.
 XX
 KW Leptin; LIG46; LIG56; leptin induced gene; Tgtp; LRG-47; RC10-II; Strai3;
 KW anorectic; anabolic; antisense therapy; mouse.
 XX
 OS Mus sp.
 XX US2001024808-A1.
 PN
 XX 27-SEP-2001.
 PD
 XX 12-MAR-2001; 2001US-0804357.
 PF

CC used in the diagnosis and treatment of pancreatitis and conditions
CC that cause abnormal hypertrophy of the heart, such as hypertension,
CC myocardial infarction, valve disease and cardiomyopathy. The
CC products can also be used in detection and cell culturing.

XX
SQ Sequence 397 AA;

Query Match 87.5%; Score 1862; DB 19; Length 397;
Best Local Similarity 86.9%; Pred. No. 1.9e-195;
Matches 345; Conservative 19; Mismatches 33; Indels 0; Gaps 0;

QY 1 MSVGRRRVLLGLIMMANVFIYLVSVSKNSQDKNGKGVIIIPKEKFWKPPSIPRAYWN 60
DB 1 MSVGRRRVLLGLIMMANVFIYLVSVSKNSQDKNGKGVIIIPKEKFWKPPSIPRAYWN 60

QY 61 RQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDNNLPDRFKDFL 120
DB 61 RQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDNNLPDRFKDFL 120

QY 121 LYLCRNYSLLIDOPKCAKPPFLLLAIAKSLIPHFARQAIRESGRETNGVNTVVRVF 180
DB 121 LYLCRNYSLLIDOPKCAKPPFLLLAIAKSLIPHFARQAIRESGRETNGVNTVVRVF 180

QY 181 LLGKTPPEDNHPDLSMDLKFESEKHQDILMNNYRTFFNLKSLKEVLFRLWVSTSCPDADF 240
DB 181 LLGKTPPEDNHPDLSMDLKFESEKHQDILMNNYRTFFNLKSLKEVLFRLWVSTSCPDADF 240

QY 241 VFKGDDVFNTHHILNLYNSLSKSKAKDLFGDVHINAGPHRDKLKYIPEVYFTGVY 300
DB 241 VFKGDDVFNTHHILNLYNSLSKSKAKDLFGDVHINAGPHRDKLKYIPEVYFTGVY 300

QY 301 PPYAGGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360
DB 301 PPYAGGGGFLYSGHLLRLYHITDQVHLYPIDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360

QY 361 KKKNNICSVDLMVLSHRKPOEMIDWISQLOSPNLK 397
DB 361 KKKNNICSVDLMVLSHRKPOEMIDWISQLOSAHLK 397

RESULT 9
AAV84683
ID AAV84683 standard; Protein; 397 AA.
XX
AC AAV84683;
XX
DT 25-JUL-2000 (first entry)
XX
DE Amino acid sequence of a human LIG46 polypeptide.
XX
KW Leptin: LIG46; body weight; leptin inducible gene; obesity; cachexia;
KW LIG56; Tgtp; LRG-47; RC10-II; Stral3.
XX
OS Homo sapiens.
XX
PN W0200015826-A2.
XX
PD 23-MAR-2000.
XX
PF 10-SEP-1999; 99WO-US20722.
XX
PR 10-SEP-1998; 98US-0150857.
PR 29-OCT-1998; 98US-0106378.
PR 19-NOV-1998; 98US-0195896.
PR 13-APR-1999; 99US-0292228.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI White D, Zhou J, Tartaglia LA;
XX
DR WPI; 2000-271461/23.
DR N-PSDB; AAA12716.
XX

PT Method for determining compounds which modulate body weight and can be
PT used to treat e.g. obesity comprises measuring the activity of leptin
PT inducible genes -
XX
PS Disclosure; Fig 8; 123pp; English.
XX
CC The present sequence represents a leptin induced LIG46 polypeptide. The
CC specification describes a method for determining whether a compound can
CC be used to modulate body weight by measuring the activity of leptin
CC inducible genes, such as LIG46. The method can be used to specifically
CC detect LIG46 and LIG56 nucleic acid molecules relative to other nucleic
CC acid molecules encoding galactosyltransferases or GTP-binding proteins.
CC They can also be used in diagnostic assays to identify the presence or
CC absence of a genetic lesion or mutation characterized by aberrant
CC modification or mis-regulation of the genes or aberrant post-
CC translational modification of the proteins. LIG46 and LIG56 proteins
CC and nucleic acid molecules can be used to treat obesity or cachexia.
CC LIG46 and LIG56 antagonists are used to treat obesity and LIG46 and
CC LIG56 agonists are used to treat low body weight. The leptin inducible
CC genes i.e. LIG46, LIG56, Tgtp, LRG-47, RC10-II and Stral3 can be used
CC to determine whether a compound modulates body weight and can then be
CC used to treat obesity or cachexia or low body weight.
XX
SQ Sequence 397 AA;

Query Match 87.5%; Score 1862; DB 21; Length 397;
Best Local Similarity 86.9%; Pred. No. 1.9e-195;
Matches 345; Conservative 19; Mismatches 33; Indels 0; Gaps 0;

QY 1 MSVGRRRVLLGLIMMANVFIYLVSVSKNSQDKNGKGVIIIPKEKFWKPPSIPRAYWN 60
DB 1 MSVGRRRVLLGLIMMANVFIYLVSVSKNSQDKNGKGVIIIPKEKFWKPPSIPRAYWN 60

QY 61 RQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDNNLPDRFKDFL 120
DB 61 RQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDNNLPDRFKDFL 120

QY 121 LYLCRNYSLLIDOPKCAKPPFLLLAIAKSLIPHFARQAIRESGRETNGVNTVVRVF 180
DB 121 LYLCRNYSLLIDOPKCAKPPFLLLAIAKSLIPHFARQAIRESGRETNGVNTVVRVF 180

QY 181 LLGKTPPEDNHPDLSMDLKFESEKHQDILMNNYRTFFNLKSLKEVLFRLWVSTSCPDADF 240
DB 181 LLGKTPPEDNHPDLSMDLKFESEKHQDILMNNYRTFFNLKSLKEVLFRLWVSTSCPDADF 240

QY 241 VFKGDDVFNTHHILNLYNSLSKSKAKDLFGDVHINAGPHRDKLKYIPEVYFTGVY 300
DB 241 VFKGDDVFNTHHILNLYNSLSKSKAKDLFGDVHINAGPHRDKLKYIPEVYFTGVY 300

QY 301 PPYAGGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360
DB 301 PPYAGGGGFLYSGHLLRLYHITDQVHLYPIDDVYTGMCLOKLGVLPEKHKGFRTDIEE 360

QY 361 KKKNNICSVDLMVLSHRKPOEMIDWISQLOSPNLK 397
DB 361 KKKNNICSVDLMVLSHRKPOEMIDWISQLOSAHLK 397

RESULT 10
AAG66118
ID AAG66118 standard; Protein; 397 AA.
XX
AC AAG66118;
XX
DT 13-MAR-2002 (first entry)
XX
DE Human LIG46 polypeptide.
XX
KW Leptin: LIG46; LIG56; leptin induced gene; Tgtp; LRG-47; RC10-II; Stral3;
KW anorectic; anabolic; antisense therapy; human.
XX
OS Homo sapiens.
XX

FH Key Location/Qualifiers
FT Peptide 1..32
FT Protein /note= "signal peptide"
FT Protein 33..397
FT Protein /note= "mature protein"
XX US2001024808-A1.
XX 27-SEP-2001.
XX 12-MAR-2001; 2001US-0804357.
XX 29-OCT-1998; 98US-106378P.
XX 19-NOV-1998; 98US-0195896.
XX 10-SEP-1998; 98US-0150837.
XX (MILL-) MILLENNIUM PHARM INC.
XX White D, Zhou J, Tartaglia LA;
XX WPI: 2001-624963/72.
XX N-PSDB; AA167869.
XX An isolated polypeptide useful for treating a weight disorder and for
PT screening compounds that may inhibit protein activity comprises a
PT leptin-induced protein.
XX Example 2; Fig 8; 46pp; English.
XX The invention relates to genes whose expression are induced by leptin.
CC Lig46 and Lig56 are novel leptin induced genes (LIG), while four other
CC genes such as Tgtp, RC10-11 and Stral3 have been previously
CC identified. The leptin induced proteins can be expressed by standard
CC recombinant methodology. The proteins and encoding polynucleotides may
CC be used in screening assays to identify compounds that may bind to it.
CC Administering a molecule, e.g., an antisense molecule, which reduces
CC expression of activity of protein selected from LIG46, LIG56, Tgtp,
CC LRP-47, RC10-11, and Stral3 may be used to treat a weight disorder. The
CC method may further comprise administering leptin. The present sequence
CC represents a human Lig46 protein.
XX Sequence 397 AA:
Query Match 87.5%; Score 1862; DB 22; Length 397;
Best Local Similarity 86.9%; Pred. No. 1.9e-195;
Matches 345; Conservative 19; Mismatches 33; Indels 0; Gaps 0;
QY 1 MSVGRRRVKKLLGILMMANVFYILVEVSKNSODKNGKGVIIIPKEFKWPPSTPRAYWN 60
DB 1 MSVGRRRIRKLLGILMMANVFYIFIMEVSKSSQEKNGKGVIIIPKEFKWIKSTPPPEAYWN 60
QY 61 REQEKLNRYNPILNRVANQTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFKDF 120
DB 61 REQEKLNRYNPILNMQTGEAGRLSNISHLNVCPEPLRVTSVVTGFNNLPDRFKDF 120
QY 121 LYLRCRNYSLIDQPKCAKPFLLAITSKLIPIHFAQQATRESWGRETNGVTVYRVF 180
DB 121 LYLRCRNYSLIDQPKCAKPFLLAITSKLIPIHFAQQATRESWGRETNGVTVYRVF 180
QY 181 LLGKTPEDNHPDLSMLKFSDEKHQDILMNYRDTFFNLKVELFLRWYVSTSCPAEF 240
DB 181 LLGQTPEDNHPDLSMLKFSDEKHQDILMNYRDTFFNLKVELFLRWYVSTSCPTFE 240
QY 241 VFKGDDVFNTHHILNLYNSLSKAKDLFTGDIHINAGPHRDKLKYIPEYFYGVY 300
DB 241 VFKGDDVFNTHHILNLYNSLSKAKDLFTGDIHINAGPHRDKLKYIPEYVYGLY 300
QY 301 PPVAGGGGLYSGPALLRLYTSRVHLYPIDDYVTGMCLOKGLVPEKHKGFRIFDIE 360
DB 301 PPVAGGGGLYSGHALRLYHTDQVHLYPIDDYVTGMCLOKGLVPEKHKGFRIFDIE 360
QY 361 KKNKNTCSYIDLMLVHSRKPQEMIDIWSOLQSAHLKC 397

DB 361 KKNKNTCSYVDLMLVHSRKPQEMIDIWSOLQSAHLKC 397
RESULT 11
AAE29093
ID AAE29093 standard; Protein; 397 AA.
XX AAE29093;
XX 27-JAN-2003 (first entry)
XX Human LIG46 protein.
XX LIG46: leptin; low body weight; chromosomal mapping; tissue typing;
KW forensic biology; transgenic; gene therapy; antianorectic; human;
KW chromosome 2.
XX Homo sapiens.
XX Key Location/Qualifiers
FT Peptide 1..32
FT Protein 33..397
FT Protein /note= "Human mature LIG46 protein"
XX WO200274905-A2.
XX 26-SEP-2002.
XX 20-NOV-2001; 2001WO-USA3345.
XX 21-NOV-2000; 2000US-0717778.
XX (MILL-) MILLENNIUM PHARM INC.
XX White DW, Zhou J, Tartaglia LA, Stricker-Krongrad A, Clausen R;
XX WPI: 2002-759886/82.
XX N-PSDB; AAD46673.
XX New isolated nucleic acid molecules, designated as leptin-induced genes
PT 46 (LIG46), useful for treating a subject having a disorder
PT characterized by undesirable level of LIG46 expression or activity,
PT such as low body weight.
XX Claim 22; Fig 6; 90pp; English.
XX The invention relates to LIG46, a gene whose expression is induced by
CC leptin. LIG46 DNA and protein are useful in treating a subject having
CC a disorder characterised by undesirable level of LIG46 expression or
CC activity, such as low body weight. They are also useful in a screening
CC assay, chromosomal mapping, tissue typing and forensic biology. The
CC probes based on the LIG46 nucleotide sequence are useful for detecting
CC transcripts or genomic sequences encoding the same or related proteins.
CC The LIG46 polypeptides are useful as immunogens for raising anti-LIG46
CC antibodies. The host cells are useful for producing non-human transgenic
CC animals. LIG46 DNA is used in gene therapy. The present sequence is
CC human LIG46 protein. LIG46 gene is located at chromosome 2.
XX Sequence 397 AA:
Query Match 87.5%; Score 1862; DB 23; Length 397;
Best Local Similarity 86.9%; Pred. No. 1.9e-195;
Matches 345; Conservative 19; Mismatches 33; Indels 0; Gaps 0;
QY 1 MSVGRRRVKKLLGILMMANVFYILVEVSKNSODKNGKGVIIIPKEFKWPPSTPRAYWN 60
DB 1 MSVGRRRIRKLLGILMMANVFYIFIMEVSKSSQEKNGKGVIIIPKEFKWIKSTPPPEAYWN 60
QY 61 REQEKLNRYNPILNRVANQTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFKDF 120
DB 61 REQEKLNRYNPILNMQTGEAGRLSNISHLNVCPEPLRVTSVVTGFNNLPDRFKDF 120

QY 121 LYLRCRNYSLIDOPKCAKPFLLAIAKSLIPHFARQAIRESWGRETNGVQTYYRVE 180
Db 121 LYLRCRNYSLIDOPKCAKPFLLAIAKSLIPHFARQAIRESWGQESNAGNQTYYRVF 180
QY 181 LGKTPPEDNHPDLSMDLKFESDKHODILMNNYRDTFFNLSKEVLFRLWVSTSCPDSEF 240
Db 181 LGKTPPEDNHPDLSMDLKFESDKHODILMNNYRDTFFNLSKEVLFRLWVSTSCPDSEF 240
QY 241 VFKGDDVFNTHILNLYNSLSKAKOLFIDGVTHNAGPHRDKKLYIPEVFTYGVY 300
Db 241 VFKGDDVFNTHILNLYNSLSKAKOLFIDGVTHNAGPHRDKKLYIPEVFTYGVY 300
QY 301 PPVAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKGLVPEKHKGFTFDEE 360
Db 301 PPVAGGGFLYSGHLLRLXHTDQVHLYPIDDVYTGMCLOKGLVPEKHKGFTFDEE 360
QY 361 KNKNKICSYIDLMLVHRSRKPQEMIDINSQLOSPNLKC 397
Db 361 KNKNKICSYVDLMLVHRSRKPQEMIDINSQLOSAHLKC 397

RESULT 12
ABB82649
ID ABB82649 standard; Protein: 397 AA.
XX AC ABB82649;
XX AC ABB82649;
DT 19-FEB-2003 (first entry)
XX Human beta3GnT polypeptide sequence.
XX Beta3GnT; transgenic; beta-1,3-N-acetylglucosaminyltransferase;
KW disease system; human; enzyme.
XX Homo sapiens.
XX WO200279413-A2.
XX 10-OCT-2002.
XX 29-MAR-2002; 2002WO-US09645.
XX 29-MAR-2001; 2001US-280706P.
PR 28-MAR-2002; 2002US-0280706.
XX (DELTA-) DELTAGEN INC.
XX Leviten MW, Phillips R;
XX WPI: 2003-067437/06.
DR N-PSDB: ABV75082.
XX
PT New transgenic mouse comprising a disruption in a beta3GnT gene, as in
PT vivo model to study various disease states or conditions in which
PT beta3GnT may be implicated or involved, such as abnormal cell growth,
PT cancer and metastasis.
XX
PS Disclosure; Fig 4: 55pp; English.
XX
CC The invention relates to a transgenic mouse comprising a disruption in a
CC beta3GnT (beta-1,3-N-acetylglucosaminyltransferase) gene, where there is
CC no native expression of beta3GnT gene. The transgenic mice may be used
CC as in vivo model to study various disease states or conditions in which
CC beta3GnT may be implicated or involved, such as abnormal cell growth,
CC cancer and metastasis, and to evaluate various treatments or to identify
CC agents for treating disease states or conditions, such as anxiety or
CC depression. Animal-based disease systems may be used to identify
CC compounds capable of ameliorating disease symptoms, as test substrates
CC for the identification of drugs, pharmaceuticals, therapies and
CC interventions that may be effective in treating a disease or other
CC phenotypic characteristic of the animal. The present sequence represents
CC a human beta3GnT polypeptide sequence.
XX

SQ Sequence 397 AA;
Query Match 87.5%; Score 1862; DB 24; Length 397;
Best Local Similarity 86.9%; Pred. No. 1.9e-195;
Matches 345; Conservative 19; Mismatches 33; Indels 0; Gaps 0;

QY 1 MSVGRRRVKKLGILMANVFLYLIVEYKSSODKNGKGVIIIPKEKFWKPPSTPRAYWN 60
Db 1 MSVGRRRRIKLGILMANVFIYIMEVSKSSQKNGKGEVILIPKEKFWKISTPEEAYWN 60
QY 61 REQEKLNRYNPIILNRVANOTGELATSPNTSHLSYCEPDSVTMTAVTFDNNLPDRFKDFL 120
Db 61 REQEKLNRYNPIILSMLTNOTGEAGRLSNISHLAYCEPDLRVTSVVTGFNNLPDRFKDFL 120
QY 121 LYLRCRNYSLIDOPKCAKPFLLAIAKSLIPHFARQAIRESWGRETNGVQTYYRVF 180
Db 121 LYLRCRNYSLIDOPKCAKPFLLAIAKSLIPHFARQAIRESWGQESNAGNQTYYRVF 180
QY 181 LGKTPPEDNHPDLSMDLKFESDKHODILMNNYRDTFFNLSKEVLFRLWVSTSCPDSEF 240
Db 181 LGKTPPEDNHPDLSMDLKFESDKHODILMNNYRDTFFNLSKEVLFRLWVSTSCPDSEF 240
QY 241 VFKGDDVFNTHILNLYNSLSKAKOLFIDGVTHNAGPHRDKKLYIPEVFTYGVY 300
Db 241 VFKGDDVFNTHILNLYNSLSKAKOLFIDGVTHNAGPHRDKKLYIPEVFTYGVY 300
QY 301 PPVAGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKGLVPEKHKGFTFDEE 360
Db 301 PPVAGGGFLYSGHLLRLXHTDQVHLYPIDDVYTGMCLOKGLVPEKHKGFTFDEE 360
QY 361 KNKNKICSYIDLMLVHRSRKPQEMIDINSQLOSPNLKC 397
Db 361 KNKNKICSYVDLMLVHRSRKPQEMIDINSQLOSAHLKC 397

RESULT 13
AAB49748
ID AAB49748 standard; protein; 413 AA.
XX AC AAB49748;
XX AC AAB49748;
DT 17-APR-2001 (first entry)
XX Human beta 1,3-N-acetylglucosamine transferase protein G3.
DE Human beta 1,3-N-acetylglucosamine transferase protein G3.
XX Sugar chain synthesis agent; beta 1,3-N-acetylglucosamine transferase;
KW inflammation; cancer; metastasis; human.
XX Homo sapiens.
XX OS Homo sapiens.
XX WO200100848-A1.
XX 04-JAN-2001.
XX 29-JUN-2000; 2000WO-JP04304.
PF 29-JUN-1999; 99JP-0183437.
PR 16-MAR-2000; 2000JP-0074757.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Sasaki K, Shiraishi N, Natsume A, Yamada Y, Nakagawa S, Sekine S;
XX WPI: 2001-102895/11.
DR N-PSDB: AAF29255.
XX
XX New polypeptide having beta1,3-N-acetylglucosamine transferase activity
PT for diagnosis of inflammation, cancer and cancer metastasis,
PT development of remedies, and for producing glycoconjugates.
XX
XX Claim 1: Page 147-150; 195pp; Japanese.
XX
CC This invention relates to a sugar chain synthesizing agent that contains

CC a polypeptide as the active ingredient, where the polypeptide has beta
 CC 1,3-N-acetylglucosamine transferase activity. The polypeptides, DNA
 CC sequences encoding them and antibodies directed against the proteins are
 CC useful in the diagnosis of inflammation, cancer and its metastasis,
 CC development of remedies, and for producing sugar chains and
 CC glycoconjugates. Sequences AAB49748 - AAB49751 represent polypeptides of
 CC the invention, having beta 1,3-N-acetylglucosamine transferase activity,
 CC and polynucleotides AAF29255 - AAF29258 represent cDNA encoding them. pcr
 CC primers used in the isolation and characterisation of the cDNA sequences
 CC are represented by sequences AAF29259 - AAF29260.
 XX
 XX Sequence 413 AA:

Query Match 86.6%; Score 1844; DB 22; Length 413;
 Best Local Similarity 83.5%; Pred. No. 1.9e-193;
 Matches 345; Conservative 19; Mismatches 33; Indels 16; Gaps 1;
 QY 1 MSVGRPRVKLLGILMM-----ANVFYILIVEVSKNSQDKNGKGVIIIP 44
 DB 1 MSVGRPRVKLLGILMMANVFYIMEVSKNSANVFYIMEVSKNSQDKNGKGVIIIP 60
 QY 45 KEFKWKPSTPRAYWNREQELKNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMT 104
 DB 61 KEFKWKPSTPRAYWNREQELKNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMT 120
 QY 105 AVTFNMLPDRKDFLLVLCRNYSLLDQPKKCAKPFLLATKSLIPHFARQAIRES 164
 DB 121 VTGFMNLPDRKDFLLVLCRNYSLLDQPKKCAKPFLLATKSLIPHFARQAIRES 180
 QY 165 WQRETNVGNQTVRVVFLGKTPEDNHPDLSMDLKFESDKHQDILMANNYRDTFFNLSKE 224
 DB 181 WQESNAGNQTVRVVFLGKTPEDNHPDLSMDLKFESDKHQDILMANNYRDTFFNLSKE 240
 QY 225 VLFLRWVSTSCPDTEFEVKGDDVFNTHHILNLSKSKAKDLFTGDIHNAHPHD 284
 DB 241 VLFLRWVSTSCPDTEFEVKGDDVFNTHHILNLSKSKAKDLFTGDIHNAHPHD 300
 QY 285 KKLKYYIVPVYGYVPPYAGGGFLYSGPALLRLYSATSRVHLYPIDDDVYTGMCQKLG 344
 DB 301 KKLKYYIVPVYGYVPPYAGGGFLYSGHALLRLYHITDQVHLYPIDDDVYTGMCQKLG 360
 QY 345 LVPEKHKGRTFDIEKNKKNICSVID:MLVHSRKPQEMIDIWLOSPNLKC 397
 DB 361 LVPEKHKGRTFDIEKNKKNICSVID:MLVHSRKPQEMIDIWLOSPNLKC 413

RESULT 14

AAV69698

ID AAV69698 standard; Protein: 397 AA.

AC AAV69698:

XX 08-MAY-2000 (first entry)

DT Human Brainiac protein.

DE Human Brainiac protein.

XX Brainiac: human; expressed sequence tag; EST;

KW Drosophila melanogaster; Egghead; Notch; epithelial cell

KW Fringe family; cell viability; growth regulation; cell fate; cancer;

KW psoriasis; skin lesion; nervous system disorder; developmental syndrome;

KW transgenic animal; drug screening.

XX Homo sapiens.

OS Homo sapiens.

XX Key

FH Location/Qualifiers

FT Misc-difference 104

FT /note- "Encoded by NNNNNN"

XX CA2255109-A1.

XX 17-JUN-1999.

XX 17-DEC-1998; 98CA-2255109.

XX

XX

XX

XX

XX

XX

XX

XX 17-DEC-1997; 97CA-2225126.
 XX (HSCR-) HSC RES & DEV LP.
 PA Egan SE;
 PI WPI; 2000-148082/14.
 DR N-PSDB; AAZ87186.
 XX New nucleic acids encoding a murine and human Brainiac protein, useful
 PT for detecting somatic or germline DNA lesions which are responsible for
 PT developmental syndromes or diseases including cancer -
 PS Claim 13; Page31; 40pp; English.
 XX This sequence represents human Brainiac protein. The complete murine
 CC Brainiac cDNA sequence (AAZ87185) was used to screen an EST (expressed
 CC sequence tag) database to identify human Brainiac cDNA clones,
 CC which were assembled to form the human Brainiac cDNA of the present
 CC invention. Human Brainiac has significant similarity to Drosophila
 CC melanogaster Brainiac. Drosophila Brainiac and Egghead proteins regulate
 CC adhesion between epithelial cells, this activity being dependent on the
 CC presence of Notch. Drosophila Brainiac is a secreted protein, and has
 CC sequence similarities with the Drosophila Fringe proteins. Brainiac/
 CC Egghead-mediated epithelial cell adhesion is required for cell
 CC viability, cell growth regulation and cell fate specification. Wild-type
 CC or mutant forms of mammalian Brainiac proteins may therefore be used to
 CC alter epithelial cell adhesion in a mammal. Mammalian Brainiac proteins,
 CC active fragments analogues, and nucleic acids may be used to treat
 CC diseases such as cancer, psoriasis and other skin lesions, and nervous
 CC system disorders. Mammalian Brainiac nucleic acids may also be used to
 CC detect somatic or germline DNA lesions which are responsible for
 CC developmental syndromes or diseases including cancer. The mammalian
 CC Brainiac proteins and fragments or its analogues are useful as antigens
 CC in immunoassays including enzyme-linked immunosorbent assays (ELISA),
 CC radioimmunoassays (RIA) and other non-enzyme linked antibody binding
 CC assays. Non-human transgenic animals comprising nucleotide sequences
 CC encoding human Brainiac protein (AAV69698) can be used as animal models
 CC for the study of mammalian Brainiac gene function, for the screening of
 CC candidate compounds and for the evaluation of potential therapeutic
 CC interventions.
 XX Sequence 397 AA;
 SQ
 Query Match 83.8%; Score 1784; DB 21; Length 397;
 Best Local Similarity 85.6%; Pred. No. 7.3e-187;
 Matches 332; Conservative 17; Mismatches 39; Indels 0; Gaps 0;
 QY 10 LIGILMANVYILIVEVSKNSQDKNGKGVIIIPKEFKWKPSTPRAYWNREQELNRW 69
 DB 10 LIGILMANVYILIVEVSKNSQDKNGKGVIIIPKEFKWKPSTPRAYWNREQELNRQ 69
 QY 70 YNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTVDNFNLDPDRKDFLLVLCRNY 125
 DB 70 YNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTVDNFNLDPDRKDFLLVLCRNY 129
 QY 130 LLIDQPKKCAKPFLLAIAKSLIPHFARQAIRESWGRNVTGTVRVVFLGKTPPED 189
 DB 130 LLIDQPKKCAKPFLLAIAKSLIPHFARQAIRESWGRNVTGTVRVVFLGKTPPED 189
 QY 190 NHPDLSMDLKFESDKHQDILMANNYRDTFFNLSLKEVLFLRWVSTSCPDTEFEVKGDDV 249
 DB 190 NHPDLSMDLKFESDKHQDILMANNYRDTFFNLSLKEVLFLRWVSTSCPDTEFEVKGDDV 249
 QY 250 VNTTHILNLSKSKAKDLFIGDVIHNAHPHDKRLKYYIPEVYTYGYPYAGGGF 309
 DB 250 VNTTHILNLSKSKAKDLFIGDVIHNAHPHDKRLKYYIPEVYTYGYPYAGGGF 309
 QY 310 LYSGPALLRLYSATSRVHLYPIDDDVYTGMCQKLGVLPEKHKGRTFDIEKNKKNICS 369
 DB 310 LYSGPALLRLYSATSRVHLYPIDDDVYTGMCQKLGVLPEKHKGRTFDIEKNKKNICS 369

QY 370 IDMLVHSRKPEMIDWSQLSPNLKC 397
:
D6 370 VDMLVHSRKPEMIDWSQLSAHLKC 397

RESULT 15

K2501 12
AAB93536

AA393536
ID AA393536 standard: protein: 373 AA.

AC ABQ3536-

XX
DT
76 - TRIN - 20

C
 D
 E
 F
 G
 H
 I
 J
 K
 L
 M
 N
 O
 P
 Q
 R
 S
 T
 U
 V
 W
 X
 Y
 Z

XX

XX

XX

XX

XX
XX
TCTC

PF
yy
Z8-JUL-2000; ZC00EP-0116128.

PR 29-JUL-1999; 99JP-0248036.

PR 11-JAN-2000; 2000JP-0118775.

PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isoqai T, Nishikawa

[illegible]

DK WPI; 2001-310749/34.
XX

PT primer sets for synt

PT and/or diagnosis of

XX
PS Claim 8; SEQ ID 12898; 2537pp + CD ROM; English.

CC The present invention describes primer sets for

comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 373 AA:

Query Match	Q938	C0000	D030	F0004	J750

Best Local Similarity 86.68; Pred. No. 1.7e-183;

Qy	25	VEVSKSSODKNGKCGVILIPKEKFWKPPSTPRATYNREQEKLNRWYNPILNRVANOTGEL	84
Db	1	MEYKSSSOEKNKGGEVILIPKEKFWKISTPPEAYWNREQEKLNRQYNPILSMLTNOITGEA	60
Qy	85	ATSPNTSHLSYCEPDSTVYATVTFNNLPDRFKDFLLYLRCRNYSLILIDQPKKCAKKPEL	144
Db	61	GRLSNISHUNYCEPDRTSVYTFGNLPDRFKDFLLYLRCRNYSLIDQPKCAKKPEL	120
Qy	145	LLAIKSLIPHFARRQAIRESNGRETNVGNQTVVRVFLLGKTPPEDNHPDLSMDLKFESEK	204
Db	121	LLAIKSLTPHFARRQAIRESNQGESNAGNQTVVRVFLLGQTPPEDNHPDLSMDLKFESEK	180
Qy	205	HQDILMWNYRDTFFNLSEKVEFLRWSTSCPDABEFVKGDGDDVFVNTHHILNLSLSEK	264
Db	181	HQDILMWNYRDTFFNLSEKVEFLRWSTSCPDTEFVKGDGDDVFVNTHHILNLSLSEK	240
Qy	265	SKAKDLFIGDVTHNAGPHRDKKLKYYIPEVFTYTCVPPYAGGGGFLYSGPALLURLYSATS	324
Db	241	TRAKDLFIGDVTHNAGPHRDKKLKYYIPEVYVYSGLYPPYAGGGGFLYSGHLAURLYHITD	300
Qy	325	RVHLYPIDDVTYGMCLQKLGVLPEKHKGFRTDIEEKNKNKICSYIDLMVLVHSRKKPQEMI	384
Db	301	QVHLYPIDDVTYGMCLQKLGVLPEKHKGFRTDIEEKNKNKICSYVDLMVLVHSRKKPQEMI	360
Qy	385	DIWSLOQSPNLKC	397
Db	361	DIWSLOQSAHLKC	373

Search completed: October 20, 2003, 23:01:58

Job time : 40.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 23:00:41 ; Search time 12.5 Seconds
(without alignments)
1343.793 Million cell updates/sec

Title: US-09-804-357b-14

Perfect score: 2123
Sequence: 1 MSVGRRIKLLGILMANVF.....RKPOEMDIWSQLQSAHLKC 397

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	616.5	29.0	378	4	US-09-482-180A-2
2	557	26.2	397	4	US-09-459-133-2
3	549	25.9	389	4	US-09-459-133-13
4	479.5	22.6	326	2	US-09-055-097-6
5	328	15.4	325	2	US-09-055-097-5
6	309.5	14.6	331	4	US-09-996-243-209
7	280	13.2	378	1	US-09-055-097-1
8	140.5	6.6	372	1	US-08-207-904-10
9	94.5	4.5	309	4	US-09-134-001C-5667
10	94	4.4	406	4	US-09-134-001C-3544
11	92.5	4.4	523	1	US-08-073-383-2
12	92.5	4.4	523	5	PCT-US94-06365-2
13	91.5	4.3	523	1	US-08-428-415-2
14	91.5	4.3	523	1	US-08-379-685-2
15	91.5	4.3	523	2	US-08-854-029-2
16	91.5	4.3	523	3	US-08-428-762-2
17	90	4.2	350	4	US-09-484-035A-11
18	90	4.2	363	4	US-09-464-035A-1
19	90	4.2	363	4	US-09-849-562A-1
20	90	4.2	363	4	US-09-849-031A-1
21	89.5	4.2	354	4	US-09-574-377-24
22	87	4.1	354	4	US-09-574-377-32
23	85.5	4.0	520	4	US-09-752-165-1
24	85	4.0	624	2	US-08-756-317-9
25	84.5	4.0	734	4	US-09-585-858-9
26	83	3.9	354	4	US-09-574-377-25
27	83	3.9	707	4	US-09-228-986-80

ALIGNMENTS

RESULT :

US-09-482-180A-2
: Sequence 2, Application US/09482180A
: Patent No. 6361985
: GENERAL INFORMATION:
: APPLICANT: Conklin, Darrell C.
: APPLICANT: Yamamoto, Gayle
: APPLICANT: Gao, Zeren
: APPLICANT: Whitmore, Theodore E.
: APPLICANT: Jaspers, Stephen
: TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING
: TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOG, ZNPS6
: FILE REFERENCE: 98-80
: CURRENT APPLICATION NUMBER: US/09/482,180A
: PRIOR FILING DATE: 2000-01-12
: PRIOR APPLICATION NUMBER: 60/115,721
: PRIOR FILING DATE: 1999-01-12
: NUMBER OF SEQ ID NOS: 20
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 378
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-482-180A-2

Query Match 29.0%; Score 616.5; DB 4; Length 378;
Best Local Similarity 43.2%; Pred. No. 4.2e-60;
Matches 133; Conservative 49; Mismatches 107; Indels 19; Gaps 8;

QY	96	CEPDLRVTSWTGTGNNLPDRFKDFLLYLRCRNVSLIDQPKCAKKPFLLLAIKSLTPHF	155
DB	76	CPNHTVSSASL---SLPSRHRFLTYRCHRFNSILL-EPSCSKDTELLAIAKSQPGHV	131
QY	156	ARQAIRESWGQESN-AGNQTVVVRVFL---GQTPPENHPDLSMLKFESEKHODILMW	211
DB	132	ERRAAIRSTWGRVGGWARGQLKLVLLGVAGSAPP-----AOLLAYESREFDDILQW	184
QY	212	NYRDTFFNLSEKLVFLRWVSTSCPTDTEPVFGKDDVFNTHHILNLYNLSLTKAKDLF	271
DB	185	DETEDFNLTLELHLQWRVVAACQAHFMLKDDVVFVHPNVLEFLDGWD--FAQDLL	242
QY	272	IGDVTHNAGPHRDKKLYKYPVWVYSGL-YPPYAGGGFLYSGHLALRYHTDQVHLYP	330
DB	243	VGDTVQALPNTNKKYFIPPSMVRATHYPYAGGGVMSRATVRRLQIMDAEELS	302
QY	331	IDDVYTCMLQKLGVLPEKHKGFRTDIEEK-KNNICSYVDLMVLVHRSKPKQEMIDWSQ	389
DB	303	IDDVFGVCMRLRLGLSPMHAGFKTIGIRRLDPLDPLCYRLGLLVHRLSPLEMTMAL	362
QY	390	LQSAHLKC	397

```

;
; TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS
; FILE REFERENCE: 98-77
; CURRENT APPLICATION NUMBER: US/09/459,133
; CURRENT FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: 60/111,697
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Mus musculus
; US-09-459-133-13

Query Match      25.9%; Score 549; DB 4; Length 389;
Best Local Similarity 32.5%; Pred. No. 1.5e-52;
Matches 136; Conservative 70; Mismatches 142; Indels 70; Gaps 12;

QY 8 IKLLGILMANVFIYFIMEVSKSSQKNGKGEVVIIPKEKFKWISTPP-----TPPNAEPTLTNLSA 58
DB 14 LTLGLKLV-----YIEWTSESWLKKAEPGALPSP-----TPPNAEPTLTNLSA 58
QY 56 -----EAYNREQEKLNQYNPILSMLTNQTGEAGRLSNISHLNLYCEPDLRVTSV 106
DB 59 RLGTGGLSSAYWNOQRQLG-----VLPSTDCQTW--GTVA-----ASEI 97
QY 107 TGFNNLPDRFKDFLLYLRCRNYSLIDQPD-----KCAKK--PELLLAIKSLTPHEARRQ 159
DB 98 LDFILYPOELRRFLLSAACSFPFLWLPAGEGSPVASCSDKDVYLLLAIVKSEPGHFAARQ 157
QY 160 AIRESWGQESNAGNQTYYRVVFLGQTTPEDNHPDLSMDLKFESSEKHODILMWNRYDTFFN 219
DB 158 AVRETWGSVPV-AGTRL---LFLLG-SPLGMGGPDLRLSVTWESRRYGDLLWDFLDVYPN 212
QY 220 LSLKEVLFRLWVSTSCPDTPEVFKGDDVDVFNTHHILNLYNSLSKTKAKDLFIGDVHNA 279
DB 213 RTLADLLLTWLSHHCDFNVFLQVQDDAFVHTPALLEHLQTLPTTWARSLYLGEIFTQA 272
QY 280 GPHRDKKLKYYIPEVYSGLYPPYAGGGGFLYSGHLALRLYHITDQVHLXPIDDVYTGMC 339
DB 273 KPLRKPGGPFVVPKTFEGDYFVYASGGYVYISGRAPWLLQAAARVAPFPDDVYTGFC 332
QY 340 LQKGLVPEKHKGFRTDIEKNNKNCISYVDLMVHSRKPQEMIDWISQLOSAHLKC 397
DB 333 FRALGLAPRAHPGFLTAWPAERTDP-CAVRGLLLVHPVSPQDTIWLWRHLWVPELQC 389

RESULT 4
US-09-055-097-6
; Sequence 6, Application US/09055097
; Patent No. 5955282
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: HUMAN OXIDIZED LDL RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/055,097

;
; TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS
; FILE REFERENCE: 98-77
; CURRENT APPLICATION NUMBER: US/09/459,133
; CURRENT FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: 60/111,697
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: VARIANT
; LOCATION: (137)...(137)
; FEATURE:
; OTHER INFORMATION: Xaa is Gly or Ser
; NAME/KEY: VARIANT
; LOCATION: (1)...(397)
; OTHER INFORMATION: Xaa = Any Amino Acid
; US-09-459-133-2

Query Match      26.2%; Score 557; DB 4; Length 397;
Best Local Similarity 34.5%; Pred. No. 2e-53;
Matches 121; Conservative 69; Mismatches 131; Indels 30; Gaps 8;

QY 55 PEAYNREQEKLNQYNPILSMLTNQTGEAGRLSNISHLNLYCEP-DLRVTSVVTGNNLP 113
DB 69 PFAYWNOQWRLG-----SLPSGSDSTETGG-----COAWGAAATAEIPDFASYP 112
QY 114 DREKDFLLYLRCRNYSLI-----DQPKCAKK--PELLLAIKSLTPHEARRQAIRESWG 166
DB 113 KDLRRFLLSAACRFPQWLPFGGCGXQVSSCSDTDVYPYLLAVKSEPGRFAERCAVRETWG 172
QY 167 QESNAGNQTYYRVVFLGQTTPEDNHPDLSMDLKFESSEKHODILMWNRYDTFFNLSLKEVL 226
DB 173 SPAPG----IRLLFLLG-SPVGEAGPDLSDLVASWESRRYSDDLWDFLOVPFNOTLKDLL 227
QY 227 FLRWVSTSCPDTPEVFKGDDVDVFNTHHILNLYNSLSKTKAKDLFIGDVHINAGPHRDKK 286
DB 228 LLWLGRHCPTVSVFLRAQDDAFVHTPALLAHLRALPPASARSYLGEVFTQAMPURKPG 287
QY 287 LKYIPEVYVSGLYPPYAGGGGFLYSGHLALRLYHITDQVHLXPIDDVYTGMCQKGLV 346
DB 288 GPFYVPEFEGGYPAYASGGYVYIAGRLAPWLLRAARVAPFPFEDVYTGICIRALGV 347
QY 347 PEKHKGFRTDIEKNNKNCISYVDLMVHSRKPQEMIDWISQLOSAHLKC 397
DB 348 PQAHGFLTAWPADRTADH-CAFRNLLLVRLPGLGQASIRLWKQLQDPRLOC 397

RESULT 3
US-09-459-133-13
; Sequence 13, Application US/09459133
; Patent No. 6416988
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Yamamoto, Gayle
; APPLICANT: Jaspers, Stephen R.
; APPLICANT: Gao, Zeren
; APPLICATION NUMBER: US/09/055,097
```

```

/ FILING DATE: Filed Herewith
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Cerrone, Michael C.
/ REGISTRATION NUMBER: 39,132
/ REFERENCE/DOCKET NUMBER: PF-0490 US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (650) 855-0555
/ TELEFAX: (650) 845-4166
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 326 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ LIBRARY: GenBank
/ CLONE: 2745735
/ US-09-055-097-6

Query Match 22.6%; Score 479.5; DB 2; Length 326;
Best Local Similarity 35.6%; Pred. No. 6.6e-45;
Matches 110; Conservative 67; Mismatches 111; Indels 21; Gaps 9;

QY 101 RVTSVVTCG---FNNLPDRFXDELL-YLRCR---NYSLLDIDOPDKCAKK-PFLLLAIKS 150
DB 27 RPTSYTSKSPFSLTVARNFTGNTRPINFHSFEFLINEPNKCEKNIPFLVILIST 86
QY 151 LTPFARQAIRAESWGQESNAGNOTVYRVFLLGQTPPEDNHPDLSDLMAFSESEKHQDILM 210
DB 87 THKEFDARQAIRWTGWDENNFKGKIATLFLG---KNADPYLNMVQESQIFHDIV 142
QY 211 WNYRTPFNLSLKEVFLRWVTSCTPTEVFYFGDDVDVFNTHILNYLNSLSKTKAKDL 270
DB 143 EDFIDSYNLTFLKLMGRWYATFCRSKAKYVMKTDSDIFVNDMLIYKLLKESTKPRRY 202
QY 271 FIGDVIHNAGRHRDKLKYIPEVYV-SGLYPPYAGGGFLXSGHLALRLYHITDQVHLY 329
DB 203 FTGYVI-NGGPIRDRSRKWTAPRDLYPDSNYPFCSGTGYIFSADVAELIYKTSUHLRL 261
QY 330 PIDDYVTGMCLQKLGVLPEKHKGFRTDIEKNKNKICSYVDLMVLVHSRKPQEMIDIWSQ 389
DB 262 HLEDYVGLCKRLGIHPFQNSGFNHW---KNAYSICRYRVIIVHQISPEEMHRIWND 317
QY 390 LOS-AHLKC 397
DB 318 MSSKKHLRC 326

RESULT 5
US-09-055-097-5
; Sequence 5, Application US/09055097
; Patent No. 5955282
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: HUMAN OXIDIZED LDL RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

```

APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730P1C13
CURRENT APPLICATION NUMBER: US/09/996,243
CURRENT FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/082250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090535
PRIOR FILING DATE: 1998-06-24

```

; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

```

```

Query Match      14.6%; Score 309.5; DB 4; Length 331;
Best Local Similarity 29.5%; Pred. No. 6.5e-26;
Matches 82; Conservative 58; Mismatches 123; Indels 15; Gaps 8;

Qy 126 RYSLIDQPKCA--KKPFLLLAIAKSLTPHARRQAIRESWGQESNAGNOTVVRVFLLGQ 184
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 QDFHFLREHSNCSSQNPFLVLTSHPSDYKARQAIRVTWGEKKSWMGVEVLTFFLLGQ 120

Qy 185 -TPPEDNHPDSDMLKFESEKHODILMNYRDTFENLSLKEVFLRLVRYSTSCDPTFVFK 243
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 EAEREDKMLAS--LEDEHLVIGDIROFDLDTNNLTKTIMAFRWVTECPNKKVYMK 178

Qy 244 GDDVVFVNTHHILNYSLSKTKAKDLFIGDVIHNAAGHRDKKLYIPEVYVS-GLYPP 302
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 TDTDFVINTGNLVKYL--LNLHSEKFTGYPLDNTSYRGYQKTHISYQYEPFKVFP 236

Qy 303 YAGGGGFLYSGHLARLYHTDQVHLPIDDVYTCMLQKLGCL---VPEKHKGRTEFIE 359
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 237 YCSGLGYTMSRDLVPRYIEMGHVKPIKFDVYVIGICLNLLKVNIIHIPEDTNLFYRIH 296

Qy 360 EKNKNICSYVDLMLVHSKRKPOEMIDWSQSAHLK 397
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 297 ----LDVQLRRVIAAHGFSSEKIIITFW-QVWLRNTTC 329

```

```

RESULT 7
US-09-055-097-1
; Sequence 1, Application US/09055097
; Patent No. 5955282
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.

```

```

; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: HUMAN OXIDIZED LDL RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/055,097
; FILING DATE: Filed Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Cerrone, Michael C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0490 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 378 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: DUODNOT02
; CLONE: 1705085
; US-09-055-097-1

```

```

Query Match      13.2%; Score 280; DB 2; Length 378;
Best Local Similarity 27.0%; Pred. No. 1.6e-22;
Matches 81; Conservative 46; Mismatches 119; Indels 54; Gaps 9;

Qy 130 LLIDQPKCA--KKPFLLLAIAKSLTPHARRQAIRESWGQESNAGNOTVVRVFLLGQTP 186
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 56 LLIPNQEACSGPGCAPFPFLILVCTAPENLNQRNATRASWGLRARGLRVQTLFLGGE-- 113

Qy 187 PEDNHP-----DLSMDLKFESEKHODILMNYRDTFENLSLKEVFLRLVRYSTSCDPT 239
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 114 PNAQHPVWGSQSDLAS---ESAAOGDILQAAPQDSYRNLTILKTLGSLNWAEEKHCPMAR 169

Qy 240 VFVKDDDDVFVNTHHILNYL-----NSLSKTKAKDLFIGD 274
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 170 YVLKTDVVVNVVVELVSELVLRGGRWGQWERSTEPQREAEQGGVLSHEVPLLYLGR 229

Qy 275 VIINAGPHRDKKLYIPEVYVS---GLYPPYAGGGGFLYSGHLARLYHTDQVHLPI 331
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 230 VHRVNPRTPGGRHRSVEQWPHWTGPPPVASGTGYVLSASAVQLILKVASRAPLPL 289

Qy 332 DDVYTCMLQKLGVLPE---KHKGRFTFIEEKNKNICSYVDMLV-HSRKPOEMIDIW 387
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 290 EDVFGVSARRGGLAPTQCVKLAGATHYPIIDR-----CCYCKFLTSLRDLDPKMQEAW 343

```

```

RESULT 8
US-08-207-904-10
; Sequence 10, Application US/08207904
; Patent No. 5477002
; GENERAL INFORMATION:
; APPLICANT: Tuttle, AnnMarie

```

```

Query Match          6.6%  Score 140.5; DB 1; Length 372;
Best Local Similarity 26.6%; Pred. No. 5.9e-07;
Matches 67; Conservative 46; Mismatches 96; Indels 43; Gaps 14;

QY      141 KPFLLLAIAKSLTPHARRQAIRESWGQESN-----AGNOTVVRVFLLGQTTPEDNHPCLS 195
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db      101 KATIVVGINTAFSSRKRDLSRETWMPKGDGKLRLEKEGIVIRFVICHSATRGGVLDRA 160

QY      196 DMLKFSEKHQDILMNRYRDTFFNLSLKEVLFLRWVSTSCPDTEFVFGDDDDVFVTHH: 255
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db      161 --IDSEDAQYKDFLRLDHVEGYHELSTKIRLYFS-KAVSIWDADFYKVKDDVDHLNLGML 217

QY      256 LNTVLSLSKTKAKD-LPIGDIVIHNAGPHROKK-LKYV:PEVVYSGLYPPYAGGGGFLYSG 313
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db      218 ---ANTLAKYKSPRVYIGCM--KSGPVLQKGVRYEPE-----YMKF-GEENKYFR 265

QY      314 HLAURLYHITDQ-----VHLPIPDVVYTCMLQKLGVLPEKHKHGRTF----- 356
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db      266 HATGQIYGISRLDASYISINGILHRYANEDVLSGWL--IGLEVE-HVDFSRMCCGCTPP 322

QY      357 DIEENK-NNIC 367
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db      323 DCEWKAAGGNIC 334

RESULT 9
US-09-134-001C-5667
; Sequence 5667, Application US/091344001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C

```

```

RESULT 10
US-09-134-001c-3544
; Sequence 3544, Application US/09134001c
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: LYNN DOUCETTE-STAMM ET AL
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3544
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001c-3544

Query Match 4.4%; Score 94; DB 4; Length 406;
Best Local Similarity 21.9%; Pred. No. 0.1;
Matches 87; Conservative 57; Mismatches 151; Indels 102; Gaps 25;

Qy 42 IPKKEFW-KISTPPE-AYNRE-----QEK-LNQYNPILSMLTNQICEA 84
Db 3 LIKKLKKWLKFTKPYKHWRSGKGFMDIKAKLQIQDKGLYRELQIPQSVKQYI--- 59
Qy 85 GRLSNLSHLYNCEPDLRVTSVVTGFNNL---PDRFKDFLLYLCRCNYSLLIDQPKCAKK 141
Db 60 -YINDQSYINFTSND-----YLGIGQLEYQPNQLDFI-----KTSIHLSRSSRLVSGN 107
Qy 142 PFLLLAIR-KSLTPHFARROAIRESWGQESNAC-----NOTWVVRVLLGGTTPPEDNHPDL 194
Db 142 PFLLLAIR-KSLTPHFARROAIRESWGQESNAC-----NOTWVVRVLLGGTTPPEDNHPDL 194

```


Db 108 SWYQLEQAISEHFNFEDALIFNSGYDANLAVFNIFKNNVV-IF-----SDQONHASI 161
QY 195 SDMLKFESEKHQDILMMNYRDTFFNLSLKEVFLRWVSTSCPDTEFFVKGDVVFN-TH 253
Db 162 IDGKLSGLSVIYQHLNYYDOLESHLA-----RHINPDQKVIIVSDVSFSTNGTK 211
QY 254 HILAVNLSLTKAKDLFIGDVHINAGPHRDKKLYIIP-EVYVSGLYPPYAGGG----- 308
Db 212 ADINRLVHL-KORYNAILIIDASHLGN-----LFEYHADIDIVTSSLSKAWAGHGLILS 267
QY 309 -----FLXSGHL-ALRLXHTIDQVHL-YPIDVYTGMLQKLGVLPE--KH 350
Db 268 SKDIKDLINKGRSLIYSSSLPSYHLVFI--QVSLQHVIEDTYR---REKLNALSEYFNH 322
QY 351 KGRFTFDEEKN-----KNNIC-----SYVDLMVH 376
Db 323 QFMELFPNQPLSNTPIKNIVCDLSASAOQYDMLFEH 359

RESULT 11
US-08-073-383-2
: Sequence 2, Application US/08073383
: Patent No. 5443962
: GENERAL INFORMATION:
: APPLICANT: Draetta, Giulio
: APPLICANT: Cottarel, Guillaume
: APPLICANT: Damagnez, Veronique
: TITLE OF INVENTION: (fillin "Insert Title of Application" IASSAY AND REAGENTS FOR
: TITLE OF INVENTION: ANTI-PROLIFERATIVE AGENTS)
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD
: STREET: 60 State Street, Suite 510
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02109

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII (text)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/073,383
: FILING DATE: 19930604
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Vincent, Matthew P.
: REGISTRATION NUMBER: 36,709
: REFERENCE/DOCKET NUMBER: MII-009
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 227-7400
: TELEFAX: (617) 227-5941
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 523 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-073-383-2

Query Match 4.4%; Score 92.5; DB 1; Length 523;
Best Local Similarity 22.0%; Pred. No. 0.23;
Matches 72; Conservative 31; Mismatches 110; Indels 115; Gaps 16;
QY 71 NPILSMLTNOTG-----EAGRLSNISHLNCEPDLRVTSVVTGNNLPDRKDFLLYLR 124
Db 204 SPVTATLSDEDDGFDVLLDGNLNKEETPSCMASLWTAPLYMRTTNLDNRCK----- 256
QY 125 CRNYSLLIDQPKCAKKPFLLLAIKSLTPHPARRQAIRESWGQESNAGNOTVVRVFLGQ 184

Db 257 -----LFDSPSLCSS-----TRSVLKRPERSQEESPSTKRRKSMGSA 296
QY 185 TTPEDNHPDLSMDLKFESEKHQDILMMNYRDTFFNLSLKEVFLRWVSTSCPDTEFFVKG 244
Db 297 SPKESTNPE-----KAHETLHQ-----SLSL-----ASSP-----KG 323
QY 245 DDDVFNVTHTHILNLSLTKAKDLFIGDVHIN-AGPHRDKKLYIIPVYVSGL----- 299
Db 324 T-----IENILNDPRDLIGDFSGYLFHTVAGKHQD--LKYSISPEINASVINGRF 372
QY 300 -----YP-PYAGGGGFLYSGHLALRLXHTIDQVHLYPIDVYTGMLQKLGVL 345
Db 373 ANLIKEFVIDCRYPYIEYEGG-----HIKGAVNLHMEEEVED--FLIKRPI 416
QY 346 VPEKHKG-FRTFDIEEKNKN--NICSIV 370
Db 417 VPTDGKRVIVFHCFESSERGRPMCRYV 444

RESULT 12
PCT-US94-06365-2
: Sequence 2, Application PC/TUS9406365
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: Assay and Reagents for Identifying
: TITLE OF INVENTION: Anti-proliferative Agents
: NUMBER OF SEQUENCES: 6
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII (text)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/06365
: FILING DATE: 06-JUN-1994
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/073,383
: FILING DATE: 04-JUN-1993
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 523 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
PCT-US94-06365-2

Query Match 4.4%; Score 92.5; DB 5; Length 523;
Best Local Similarity 22.0%; Pred. No. 0.23;
Matches 72; Conservative 31; Mismatches 110; Indels 115; Gaps 16;
QY 71 NPILSMLTNOTG-----EAGRLSNISHLNCEPDLRVTSVVTGNNLPDRKDFLLYLR 124
Db 204 SPVTATLSDEDDGFDVLLDGNLNKEETPSCMASLWTAPLYMRTTNLDNRCK----- 256
QY 125 CRNYSLLIDQPKCAKKPFLLLAIKSLTPHPARRQAIRESWGQESNAGNOTVVRVFLCQ 184
Db 257 -----LFDSPSLCSS-----TRSVLKRPERSQEESPSTKRRKSMGSA 296
QY 185 TTPEDNHPDLSMDLKFESEKHQDILMMNYRDTFFNLSLKEVFLRWVSTSCPDTEFFVKG 244
Db 297 SPKESTNPE-----KAHETLHQ-----SLSL-----ASSP-----KG 323
QY 245 DDDVFNVTHTHILNLSLTKAKDLFIGDVHIN-AGPHRDKKLYIIPVYVSGL----- 299
Db 324 T-----IENILNDPRDLIGDFSGYLFHTVAGKHQD--LKYSISPEINASVINGRF 372
QY 300 -----YP-PYAGGGGFLYSGHLALRLXHTIDQVHLYPIDVYTGMLQKLGVL 345
Db 373 ANLIKEFVIDCRYPYIEYEGG-----HIKGAVNLHMEEEVED--FLIKRPI 416
QY 346 VPEKHKG-FRTFDIEEKNKN--NICSIV 370

Db 417 VPTDGRVIVVFHCFESSERGRMCRV 444

RESULT 13

US-08-428-415-2

; Sequence 2, Application US/08428415

; Patent No. 5756335

; GENERAL INFORMATION:

; APPLICANT: Cold Spring Harbor Laboratory

; TITLE OF INVENTION: No. 5756335el Human cdc25 Genes, Encoded

; TITLE OF INVENTION: Products and Uses Thereof

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: LAHIVE & COCKFIELD

; STREET: 60 State Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: ASCII(text)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/428.415

; FILING DATE: 24 April 1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Vincent, Matthew P.

; REGISTRATION NUMBER: 36,709

; REFERENCE/DOCKET NUMBER: MII-019CP3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-227-7400

; TELEFAX: 617-227-5941

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 523 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-428-415-2

Query Match 4.3%; Score 91.5; DB 1; Length 523;

Best Local Similarity 22.0%; Pred. No. 0.23;

Matches 72; Conservative 31; Mismatches 110; Indels 115; Gaps 16;

QY 71 NPILSMLTNQTG-----EAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRFKDELLYL 124

Db 204 SPVTATLSDEDDGFVDDLDGGENLKNEETPCMASLWTAPLVMRTTLDNRCK----- 256

QY 125 CRNYSLLIDQPKCAKKPFLLLAIXSLTPHFARQAIRESWGQESNAGNOTVVRVFLIGQ 184

Db 257 -----LFDSPSLCSSS-----TRSVLKRPERSQBESPPGSKRRKSMGA 296

QY 185 TPDNHPDLSMLKFESEKHQDILMNYRDTFFNLSLKEVLFRLWYSTSCPDTFEVFKG 244

Db 297 SPKESTNPE-----KAHETLHQ-----SLSL-----ASSP-----KG 323

QY 245 DDDVFVNTHHLLNLYNSLSKTKAKDLFGDVIHN-AGPHROKCLKYIPEVYVSGL----- 299

Db 324 T-----IENILDNDPRDLIGDFSKGLFHTVAGKHOD--LKYISPEIMASVINGKF 372

QY 300 -----YP-PYAGGGGFLYSGHLALRLYHITDQVHLYPIDDVYTGMCLOKLG 345

Db 373 ANLIKEFVIDCRYPYVEGG-----HIKGAVALNLMEEVED--FLKKPI 416

QY 346 VPEKHKG-FRTFDIEKNKN--NICSYV 370

Db 417 VPTDGRVIVVFHCFESSERGRMCRV 444

RESULT 14

US-08-379-685-2

; Sequence 2, Application US/08379685

; Patent No. 5770423

; GENERAL INFORMATION:

; APPLICANT: Cold Spring Harbor Laboratory

; TITLE OF INVENTION: No. 5770423el Human cdc25 Genes, Encoded

; TITLE OF INVENTION: Products and Uses Thereof

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: LAHIVE & COCKFIELD

; STREET: 60 State Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: ASCII(text)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/379.685

; FILING DATE: 26 January 1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Vincent, Matthew P.

; REGISTRATION NUMBER: 36,709

; REFERENCE/DOCKET NUMBER: MII-019-DV

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-227-7400

; TELEFAX: 617-227-5941

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 523 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-379-685-2

Query Match 4.3%; Score 91.5; DB 1; Length 523;

Best Local Similarity 22.0%; Pred. No. 0.29;

Matches 72; Conservative 31; Mismatches 110; Indels 115; Gaps 16;

QY 71 NPILSMLTNQTG-----EAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRFKDELLYL 124

Db 204 SPVTATLSDEDDGFVDDLDGGENLKNEETPCMASLWTAPLVMRTTLDNRCK----- 256

QY 125 CRNYSLLIDQPKCAKKPFLLLAIXSLTPHFARQAIRESWGQESNAGNOTVVRVFLIGQ 184

Db 257 -----LFDSPSLCSSS-----TRSVLKRPERSQBESPPGSKRRKSMGA 296

QY 185 TPDNHPDLSMLKFESEKHQDILMNYRDTFFNLSLKEVLFRLWYSTSCPDTFEVFKG 244

Db 297 SPKESTNPE-----KAHETLHQ-----SLSL-----ASSP-----KG 323

QY 245 DDDVFVNTHHLLNLYNSLSKTKAKDLFGDVIHN-AGPHROKCLKYIPEVYVSGL----- 299

Db 324 T-----IENILDNDPRDLIGDFSKGLFHTVAGKHOD--LKYISPEIMASVINGKF 372

QY 300 -----YP-PYAGGGGFLYSGHLALRLYHITDQVHLYPIDDVYTGMCLOKLG 345

Db 373 ANLIKEFVIDCRYPYVEGG-----HIKGAVALNLMEEVED--FLKKPI 416

QY 346 VPEKHKG-FRTFDIEKNKN--NICSYV 370

Db 417 VPTDGRVIVVFHCFESSERGRMCRV 444

RESULT 15

US-08-854-029-2

; Sequence 2, Application US/08854029

; Patent No. 5994074

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 22:54:34 ; Search time 12.5 Seconds
(without alignments)
1493.568 Million cell updates/sec

Title: US-09-804-357B-14
Perfect score: 2123
Sequence: 1 MSVGRRIKLLGILMANVF.....RKPEMIDIWSQLQSAHLXC 397

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2123	100.0	397	1 B3G7_HUMAN	Q9Y297 h beta-1,3-
2	601	28.3	372	1 B3G8_HUMAN	O9Y2a9 h beta-1,3-
3	418	19.7	308	1 B3G5_MOUSE	O9J167 m beta-1,3-
4	363	17.1	301	1 B3G5_PANPA	Q9n294 p beta-1,3-
5	360	17.0	300	1 B3G5_GORGO	Q9n293 g beta-1,3-
6	360	17.0	310	1 B3G5_HUMAN	Q9Y2c3 h beta-1,3-
7	358	16.9	297	1 B3G5_PANTR	Q9n295 p beta-1,3-
8	298.5	14.1	371	1 B3G4_MOUSE	Q9z0f0 mus musculus
9	298.5	14.1	371	1 B3G4_RAT	O88178 rattus norv
10	280	13.2	378	1 B3G4_HUMAN	O96024 homo sapien
11	110	5.2	1026	1 BGAL_STRTR	P23989 streptococc
12	104	4.9	1487	1 MDS3_YEAST	P53094 saccharomyc
13	102.5	4.8	453	1 YAW6_SCHPO	Q10181 schizosacch
14	102.5	4.8	544	1 YW72_METJA	Q38867 methanococc
15	95.5	4.5	351	1 YZ39_METJA	Q60294 methanococc
16	94.5	4.5	485	1 SYE2_THETIN	Q8r7t1 thermoplaer
17	93	4.4	1507	1 Y056_HUMAN	P42695 homo sapien
18	91.5	4.3	523	1 MP11_HUMAN	P30304 homo sapien
19	91.5	4.3	646	1 PIXB_HUMAN	Q14155 homo sapien
20	91	4.3	374	1 LFNG_BRARE	Q74762 brachydanio
21	90	4.2	279	1 ATPG_MYCGE	P47640 mycoplasma
22	89.5	4.2	354	1 PON3_HUMAN	Q15166 homo sapien
23	89.5	4.2	619	1 DP25_PYRAB	Q9V2f3 pyrococcus
24	89.5	4.2	1110	1 VGLM_INSV	Q01260 impatiens n
25	89.5	4.2	4036	1 RRPL_DUGBV	Q66431 dugbe virus
26	89	4.2	1916	1 RIF1_YEAST	P29539 saccharomyc
27	88.5	4.2	395	1 CGA2_CHICK	P33449 gallus gall
28	88.5	4.2	664	1 YMC0_YEAST	Q03722 saccharomyc
29	87.5	4.1	512	1 VP40_HSV7J	P52351 human herpe
30	87	4.1	343	1 YD57_METJA	Q58752 methanococc
31	87	4.1	626	1 HTPG_BUCBP	Q89a93 buchnera ap
32	87	4.1	785	1 PTAL_YEAST	Q01329 saccharomyc
33	86.5	4.1	2376	1 TAO3_YEAST	P40468 saccharomyc

RESULT 1

B3G7_HUMAN

ID B3G7_HUMAN STANDARD; PRI: 397 AA.

AC Q9N979; Q9NQ09; Q9NQ09; Q9NUT9;

DI 16-OCT-2001 (Rel. 40, Created)

DI 15-SEP-2003 (Rel. 42, Last sequence update)

DI 15-SEP-2003 (Rel. 42, Last annotation update)

DE Beta-1,3-galactosyltransferase 7 (EC 2.4.1.1.) (Beta-1,3-GalTase 7)

DE (beta3gal-17) (beta3gal-17) (UDP-galactose:beta-N-acetylglucosamine beta-1,3-galactosyltransferase 7) (UDP-Gal:beta-GlcNAc beta-1,3-galactosyltransferase 7) (Beta-3-Gx-T7).

GN B3GNT1 OR B3GALT7.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

CX MCB_TaxID=9606;

RI [1]

RP SEQUENCE FROM N.A.

RA Anado M., Carneiro F., Clausen H.;

RI "Cloning and expression of two beta-1,3-galactosyltransferases: beta3gal-T5 and beta3gal-T6.";

RI Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).

RC TISSUE=Urinary bladder;

RA Gromova I., Gromov P., Celis J.E.;

RI "A novel member of beta-1,3-galactosyltransferase family is down regulated during bladder ICC progression.";

RI Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RC TISSUE=Brain, and Testis;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Uddin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RI "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";

RI Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [4]

RP SEQUENCE OF 25-397 FROM N.A.

RN TISSUE=Placenta;

34 85.5 4.0 289 1 SYLB_AQDAE O67646 aquifex aeo

35 85 4.0 533 1 THAS_MOUSE P36423 mus musculus

36 85 4.0 1144 1 M2AL_HUMAN Q16706 homo sapien

37 84.5 4.0 696 1 SCK1_SCHPO P50530 schizosacch

38 84.5 4.0 704 1 NEUL_PIG Q02038 sus scrofa

39 84.5 4.0 1018 1 YC14_METJA Q58611 methanococc

40 84.5 4.0 1097 1 DP04_YEAST P15436 saccharomyc

41 84 4.0 432 1 BFRA_THEMA O33833 thermotoga

42 84 4.0 502 1 C83A_ARATH P48421 arabidopsis

43 84 4.0 1150 1 SCC3_YEAST P40541 saccharomyc

44 83.5 3.9 452 1 PIV2_ADE12 P12540 human adeno

45 83.5 3.9 518 1 RN23_HUMAN Q9hcm9 homo sapien

ALIGNMENTS

RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
RA Tanai H., Kinata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,
RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K.,
RA Masuko Y., Kanehori K.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP REVIEW.
RX MEDLINE=20047730; PubMed=10580128;
RA Amado M., Almeida R., Schwientek T., Clausen H.;
RT "Identification and characterization of large galactosyltransferase
RT gene families: galactosyltransferases for all functions.";
RL Biochim. Biophys. Acta 1473:35-53(1999).
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9NY97-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9NY97-2; Sequence=VSP_001791;
CC Note=NO experimental confirmation available;
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC -!- CAUTION: WAS INDICATED AS B3GAL-T6 IN SUBMITTED DNA ENTRIES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ006077; CAR91546.1;
DR EMBL; AF288208; AAF97253.1;
DR EMBL; AF288209; AAF97254.1;
DR EMBL; BC030579; AAH30579.1;
DR EMBL; BC047933; AAH47933.1;
DR EMBL; AK002009; BAA92031.1; ALT_INIT.
DR Genew; HGNC:15629; B3GNT1.
DR GO; GO:0016021; C: integral to membrane; NAS.
DR GO; GO:0008499; F: UDP-galactose beta-N-acetylglucosamine beta. .; NAS.
DR InterPro; IPR002659; Glyco_trans_31.
DR Pfam; PF01762; Galactosyl_T. 1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Signal-anchor; Golgi stack; Multigene family; Alternative splicing.
FT DOMAIN 1 7
FT TRANSMEM 8 28
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT LUMENAL, CATALYTIC (POTENTIAL).
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 79 79
FT CARBOHYD 89 89
FT CARBOHYD 127 127
FT CARBOHYD 173 173
FT CARBOHYD 219 219
FT VARSPLIC 1 11
FT MSVRRRIKLL -> MVSRLV (in isoform 2).
FT /FTID=VSP_001791.
FT CONFLICT 11 11 L -> LL (IN REF. 1).
SQ SEQUENCE 397 AA; 46022 MW; B104ECCAE26DC4AC CRC64;
Query Match 100.0%; Score 2123; DB 1; Length 397;
Best Local Similarity 100.0%; Pred. No. 8.4e-169;
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSVRRRIKLLGLMMANVFYFMEVSKSSQSKNGKNGKEVIFPKFKWISTPPEAYWN 60
Db 1 MSVRRRIKLLGLMMANVFYFMEVSKSSQSKNGKNGKEVIFPKFKWISTPPEAYWN 60
QY 61 RQEKLNQYNPILSMLTNTQGTAGRLSNISHLNVCPEDLRVTSVVTGFNNLPDRFKDFL 120
Db 61 RQEKLNQYNPILSMLTNTQGTAGRLSNISHLNVCPEDLRVTSVVTGFNNLPDRFKDFL 120

QY 121 LYLRCRNYSLIDQPKCAKPPPELLLAISLTPHFARRQAIRESWQESNAGNQTVRVF 180
Db 121 LYLRCRNYSLIDQPKCAKPPPELLLAISLTPHFARRQAIRESWQESNAGNQTVRVF 180
QY 181 LLGOTPPEDNHPDLSMDLKFSEKHDILMWNRYRDTFFNLSLKEVLFLRWVSTSCPDTF 240
Db 181 LLGOTPPEDNHPDLSMDLKFSEKHDILMWNRYRDTFFNLSLKEVLFLRWVSTSCPDTF 240
QY 241 VFKGDDVFNTHILNLSLSKTKAKDLFIDGVIHNAAGPHRDKKLKYIPEVYVSGLY 300
Db 241 VFKGDDVFNTHILNLSLSKTKAKDLFIDGVIHNAAGPHRDKKLKYIPEVYVSGLY 300
QY 301 PPYAGGGFLYSGHLARLYHIITDQVHLXPIDOVYTGMCLOKGLVPEKHKGFRTFDEE 360
Db 301 PPYAGGGFLYSGHLARLYHIITDQVHLXPIDOVYTGMCLOKGLVPEKHKGFRTFDEE 360
QY 361 KKNKNICSYVDLMLVHSRKPQEMID1WSQLQSAHLKC 397
Db 361 KKNKNICSYVDLMLVHSRKPQEMID1WSQLQSAHLKC 397
RESULT 2
B3G8_HUMAN
ID B3G8_HUMAN STANDARD; PRT; 372 AA.
AC Q9X2A9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Beta-1,3-galactosyltransferase 8 (EC 2.4.1.-) (Beta-1,3-GalTase 8)
DE (Beta3Gal-T8) (B3Gal-T8) (UDP-galactose:beta-N-acetylglucosamine beta-
DE 1,3-galactosyltransferase 8) (UDP-Gal:beta-GlcNAc beta-1,3-
DE galactosyltransferase 8) (Beta-3-Gx-T8) (Core 1 extending beta-1,3-N-
DE acetylglucosaminyltransferase) (Corel-beta3GlcNAcT).
DE B3GNT3 OR B3GALT8 OR TWEM3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99173880; PubMed=10072769;
RA Yokoyama-Kobayashi M., Yamaguchi T., Sekine S., Kato S.;
RT "Cell surface of cDNAs encoding putative type II membrane proteins on the
RT secretion from a human full-length cDNA bank.";
RL Gene 228:161-167(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21332592; PubMed=11439191;
RA Yeh J., Hiraoka N., Petryniak B., Nakayama J., Ellies L.G., Rabuka D.,
RA Hindsgaul O., Marth J.D., Lowe J.B., Fukuda M.;
RT "Novel sulfated lymphocyte homing receptors and their control by a
RT core1 extension beta1,3-N-acetylglucosaminyltransferase.";
RL Cell 105:957-969(2001).
RN [3]
RP SEQUENCE FROM N.A.
RA Jensen M.A., Bennett E.P.;
RT "Cloning of a new member of the beta 1,3 Galactosyltransferase family,
RT bl,3gal-T6.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP REVIEW.
RX MEDLINE=20047730; PubMed=10580128;
RA Amado M., Almeida R., Schwientek T., Clausen H.;
RT "Identification and characterization of large galactosyltransferase
RT gene families: galactosyltransferases for all functions.";
RL Biochim. Biophys. Acta 1473:35-53(1999).
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AB015630; BAA76497.1; -
 DR EMBL: AF293973; AAK00849.1; -
 DR EMBL: AJ130847; CAC45044.1; -
 DR Genew: H0NC13528; B3GNT3.
 DR GO: GO:0005887; C: integral to plasma membrane; TAS.
 DR InterPro: IPR002659; Glyco_trans_31.
 DR Pfam: PF01762; Galactosyl_T; 1.
 DR Transferase: Glycosyltransferase; Glycoprotein; Transmembrane;
 KW Signal-anchor; Golgi stack; Multigene family.
 FT DOMAIN 1 10 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 11 31 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 32 372 LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 64 64 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 202 202 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 362 362 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 367 367 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 372 AA: 42515 MW: 463788AFCFC2 CRC64;

Query Match 28.3%; Score 601; DB 1; Length 372;
 Best Local Similarity 41.3%; Pred. No. 1.7e-42;
 Matches 125; Conservative 51; Mismatches 117; Indels 10; Gaps 5;

QY 103 TSVVT--GFNNLPDRKDFLLYURCRNYSLID-QPKCAKPFLLAIAKSLTPFAFRQ 159
 DB 65 TSMVTHPDEATQPHVQNFLLYRCHRFLLQDVPPSKCAQPVFLLLVIKSSPSNYVRE 124
 QY 160 AIRESGQESNAGNQTWVRVLLGQTPPEDNHPDLSMDLKFSEKHQDILMWNRYDTFFN 219
 DB 125 LLRTWGREKVRGLQLRLFLVGTASNPEHAKVRLLLEAQTGDIQWDFHDSFFN 184
 QY 220 LSKEVFLRWVSTCPDTEFEVFKGDDVFNTHHILNLSKTKAKDLFGDGVHNA 279
 DB 185 LILKQVFLQWQETRCANASVFLNGDDVFAHTDNVVFYIQ--DHPDGRHLFGQLIQNY 242
 QY 280 GPHROKLLKYIPEVY-YSGLYPPYAGGGFLYSGHLALRLYHITDQVHLPIDDVYTCM 338
 DB 243 GPRAFWSKYVPEVYVQNERYPYCGGGFLLSRFTAAALRAAHVLDIFPIDVFLGM 302
 QY 339 CLQKGLVPEKHGFRFTDIEKNKN----NICSYVDLMLVHSRKPQEMIDVWSQLQSH 394
 DB 303 CLEELGLKASHSGIRTSVGRAPSOHLSSFDPCFYRDLLLVIRFLPYEMLLMMDALNQPN 362
 QY 395 LKC 357
 DB 363 LTC 365

RESULT 3
 B3G5_MOUSE
 ID B3G5_MOUSE STANDARD; PRT; 308 AA.
 AC Q9J167;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Beta-1,3-galactosyltransferase 5 (EC 2.4.1.-) (Beta-1,3-GalTase 5)
 DE (Beta3Gal-T5) (b3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
 DE 1,3-galactosyltransferase 5) (UDP-gal:beta-GlcNAc beta-1,3-
 DE galactosyltransferase 5) (Beta-3-Gal-T5) (Stage-specific embryonic
 DE antigen-3 synthase) (SSEA-3 synthase).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-129/SVJ;
 RX MEDLINE-20390006; PubMed-10837462;
 RA Zhou D., Henion T.R., Jungkawa F.B., Berger E.G., Hennet I.;
 RT "The beta1,3-galactosyltransferase beta3GalT-V is a stage-specific
 RT embryonic antigen-3 (SSEA-3) synthase";
 RL J. Biol. Chem. 275:22631-22634(2000).
 CC -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
 CC WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
 CC GLCNAC(BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOPOLIPID LC3GAL AS AN
 CC EFFICIENT ACCEPTOR. ALSO CATALYZES THE TRANSFER OF GAL TO THE
 CC TERMINAL GALNAC UNIT OF THE GLOBOSIDE GB4. THEREBY SYNTHESIZING
 CC THE GLYCOPOLIPID GB5, ALSO KNOWN AS THE STAGE-SPECIFIC EMBRYONIC
 CC ANTIGEN-3 (SSEA-3).
 CC -!- PATHWAY: Glycosylation.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
 CC -!- TISSUE SPECIFICITY: Expressed in brain and kidney.
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AF254738; AAF86241.1; -
 DR MGD; MGI:2136878; B3galT5.
 DR InterPro: IPR002659; Glyco_trans_31.
 DR Pfam: PF01762; Galactosyl_T; 1.
 KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
 KW Signal-anchor; Golgi stack; Multigene family.
 FT DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 8 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 26 308 LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 128 128 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 172 172 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 229 229 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 308 AA: 35964 MW: 789073A5178825B1 CRC64;

Query Match 19.7%; Score 418; DB 1; Length 308;
 Best Local Similarity 31.6%; Pred. No. 2.1e-27;
 Matches 91; Conservative 66; Mismatches 113; Indels 18; Gaps 7;

QY 109 FNNLPDRKDFLLYURCRNYSLIDQPD-KCAKK-PFLLAIAKSLTPHARRQAIRESWG 166
 DB 28 PRELPFVEK-----KSHGKFLQIPDIDCKQKPPFLVLLVTSKHLAARMAIRKTWG 79
 QY 167 OESNAGNOTVVRVFLGOTPPEDNHPDLSMDLKFSEKHQDILMWNRYDTFFNLSLKEVL 226
 DB 80 RETSVQGOQVTRTFLLGTSDSTEE----MDATTLESEQRDIQDKFDAYFNLTLM 135
 QY 227 FLRWVSTSCPTEFEVFKGDDVFNTHHILNLSKTKAKDLFGDGVHNAHGRHDKK 286
 DB 136 GMEVYHFCPOTAYVMKTDSDMFVNVGYLTEL--LKNKKTTFRTFGVYKPHDFPIRQKF 193
 QY 287 LKYIPEVYVS-GLYPPYAGGGFLYSGHLALRLYHITDQVHLPIDDVYTCMLQKLG 345
 DB 194 NKWFYSKFEYPWDYPPFCGTYGVSSDAIQVNVSEVPFFKLEDFVFGGLCLAKLI 253
 QY 346 VPEKHGKFRTPDIEKNKNKNCYSVDLMLVHSRKPQEMIDVWSOLOSA 393
 DB 254 RPEELHTKQTF-FPGGLRFSVCRFOKIVACHFMKPKQDILLTYWQALENS 300

RESULT 4
 B3G5_PANPA
 ID B3G5_PANPA STANDARD; PRT; 301 AA.
 AC Q9N294;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)


```

B3G5_HUMAN
ID B3G5_HUMAN STANDARD; PRT; 310 AA.
AC Q9Y2C3; Q9NY96; Q9PIX6; Q9PIX7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Beta-1,3-galactosyltransferase 5 (EC 2.4.1.-) (Beta-1,3-GalTase 5)
DE (Beta3Gal-T5) (B3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
DE 1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
DE galactosyltransferase 5) (Beta-3-Gx-T5).
GN B3GALT5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colorectal adenocarcinoma;
RX MEDLINE=99230269; PubMed=10212226;
RA Isshiki S., Togayachi A., Kudo T., Nishihara S., Watanabe M.,
RA Kubota T., Kitajima M., Shiraishi N., Sasaki K., Andoh T.,
RA Narimatsu H.;
RT "Cloning, expression, and characterization of a novel UDP-
RT galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase
RT (beta3Gal-T5) responsible for synthesis of type 1 chain in colorectal
RT and pancreatic epithelia and tumor cells derived therefrom.";
RL J. Biol. Chem. 274:12499-12507(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99337698; PubMed=10406968;
RA Zhou D., Berger E.G., Hennes T.;
RT "Molecular cloning of a human UDP-galactose:GlcNAc:beta1,3GalNAc
RT beta1,3 galactosyltransferase gene encoding an O-linked
RT core3-elongation enzyme.";
RL Eur. J. Biochem. 263:571-576(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Amado M., Carneiro F., Clausen H.;
RT "Cloning and expression of two beta-1,3-galactosyltransferases:
RT beta3Gal-T5 and beta3Gal-T6.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Tokoi Y., Choi D.-K., Groner Y.,
RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Pagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [5]
RP SEQUENCE OF 1-298 FROM N.A.
RA Liu Y., Saitou N.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP REVIEW.
RX MEDLINE=20047730; PubMed=10580128;
RA Amado M., Almeida R., Schwientek T., Clausen H.;
RT "Identification and characterization of large galactosyltransferase
RT gene families: galactosyltransferases for all functions.";
RL Biochim. Biophys. Acta 1473:35-53(1999).
CC -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
CC WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
CC GLCNAC(BETA1,3)GALNAc STRUCTURE. CAN USE GLYCOLIPID LC3GAL AS AN
CC EFFICIENT ACCEPTOR.

```

RESULT 7

B3G5_PANTR STANDARD; PRT; 297 AA.

AC QN295;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Beta-1,3-galactosyltransferase 5 (EC 2.4.1.1) (Beta-1,3-GalTase 5)

DE (Beta3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-1,3-

DE 1,3-galactosyltransferase 5) (UDP-Gal-beta-GlcNAc beta-1,3-

DE galactosyltransferase 5) (Beta-3-Gx-T5) (Fragment).

GN B3GALT5.

OS Pan troglodytes (Chimpanzee).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

OX NCBI_TaxID=9598;

RN [1]

RP SEQUENCE FROM N.A.

RA Liu Y., Saitou N.;

RA Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS

CC WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN

CC GLCNAC (BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOPOLIPID LC3CER AS AN

CC EFFICIENT ACCEPTOR (BY SIMILARITY).

CC -1- PATHWAY: Glycosylation.

CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).

CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AB041414; BAA94499.1;

DR InterPro: IPR002659; Glyco_trans_31.

DR Pfam: PF01762; Galactosyl_T; 1.

KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;

KW Signal-anchor; Golgi stack; Multigene family.

FT DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 8 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)

FT FT

FT DOMAIN 29 >297 LUMENAL, CATALYTIC (POTENTIAL).

FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT NON_TER 297 297

SQ SEQUENCE 297 AA; 34850 MW; BB7963250A637A28 CRC64;

Query Match 16.98; Score 358; DB 1; Length 297;

Best Local Similarity 29.34; Pred. No. 1.9e-22;

Matches 81; Conservative 64; Mismatches 117; Indels 14; Gaps 7;

QY 116 FKQ-FLYLRCRNYSLIDDPKCAKK--PFLLAIKSLTPHFAQALRESKGQESNAG 172

DB 31 FKQSFYKKDGNFLK---PDTDCQTPPLVLLVTSRQLAERNALQTMGKRTYK 87

QY 173 NOTVVRVLLGQTPPDNDHPLDMLKFESEKHQDILMMNYRDTFFNLSLKEVLFRLWYS 232

DB 88 GKQLKTFELLGTTSSAAETKEVDQ---ESQRGDIQKDFLDVYVNLTKTMGIEWVH 143

QY 233 TSCPDTEFVKGGDDVPVNTTHILNYLNSLTKAKDIFIGDVHINAGHRDKLKYIYP 292

DB 144 RFPQAAVFNKIDSDMFINDYLETLL--LKKNRTRFTFTGFLKLNFFPIQPFKWFVS 201

QY 293 EWTYS-GLYPYAGGGFLYSGHLAURLYHITDQVHLYPIDDYITGMCLOKGLVPEKH 351

DB 202 KSETPWDRYPPFCSTGTVFSGDVASQVYNVSEVPYKLEQVFGLCLERLNIRLEE 261

QY 352 GFRTFIDEENKNNKNSYVDMLVHVRKPKQEMIDIW 387

DB 262 SQTFP-FPGGLRFSVCFRIVACHFKPKPTLLDIW 296

RESULT 8

B3G4_MOUSE

ID B3G4_MOUSE STANDARD; PRT; 371 AA.

AC Q920F0; Q91VC1; Q920U8; Q920U9;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DE Beta-1,3-galactosyltransferase 4 (EC 2.4.1.62) (Beta-1,3-GalTase 4)

DE (Beta3Gal-T4) (UDP-galactose:beta-1,3-galactosyltransferase) (UDP-

DE galactose:beta-N-acetyl-galactosamine-beta-1,3-galactosyltransferase)

DE (GAL-T2).

GN B3GALT4.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL

RP STAGE.

RP TISSUE=Fetal;

RC MEDLINE=99432136; PubMed=10502288;

RA Daniotti J.L., Martina J.A., Zurita A.R., Maccioni H.J.F.;

RT "Mouse beta 1,3-galactosyltransferase (GAL/GMI/GD1b synthase): protein

RT characterization, tissue expression, and developmental regulation in

RT neural retina.";

RL J. Neurosci. Res. 58:318-327(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=129/SVJ;

RA Rowen L., Qin S., Madan A., Loretz C., Hall J., James R., Dors M.,

RA Shaffer T., Abbasi N., Ratcliffe A., Dickhoff R., Lasky S., Hood L.;

RT "Sequence of the mouse major histocompatibility complex class II

RT region.";

RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 1-370 FROM N.A.

RC STRAIN=BLG2/Msf, BFN/2Msf, C57BL/10SnJ, CAST/Ei, HMI/Msf, MSM/Msf,

RC NjL/Msf, pgn2, and SWN/Msf;

RA Liu Y., Kitano T., Koide T., Shirosaki T., Moriaki K., Saitou N.;

RT "Conspicuous differences among gene genealogies of 21 nuclear genes of

RT five Mus musculus subspecies.";

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: Involved in GMI/GD1B/GAL ganglioside biosynthesis.

CC -1- CATALYTIC ACTIVITY: UDP-galactose + N-acetyl-D-galactosaminyl-(N-

CC acetylneuraminyl)-D-galactosyl-D-glucosyl-N-acylsphingosine = UDP

CC + D-galactosyl-N-acetyl-D-galactosaminyl-(N-acetylneuraminyl)-D-

CC galactosyl-D-glucosyl-N-acylsphingosine.

CC -1- PATHWAY: Glycosylation.

CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.

CC -1- TISSUE SPECIFICITY: Expressed in heart, brain, spleen, kidney,

CC lung and testis.

CC -1- DEVELOPMENTAL STAGE: First expressed at embryonic day 3.

CC Maintained at high levels between days 4 and 7 and declines

CC thereafter to stabilize at low levels after day 10.

CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AF110520; AAC97977.1;

DR EMBL: AF082504; AAC69622.1;

DR EMBL: AF100956; AAC69897.1;

DR EMBL: AB039164; BAB68688.1;

DR EMBL: AB039165; BAB68689.1;

DR EMBL: AB039167; BAB68691.1;

DR EMBL: AB039166; BAB68692.1;

```

DR EMBL; AB039170; BAB68694.1; -
DR EMBL; AB039171; BAB68695.1; -
DR EMBL; AB039172; BAB68696.1; -
DR EMBL; AB039169; BAB68693.1; -
DR EMBL; AB039166; BAB68690.1; -
DR MGD; MGI:1859517; B3galT4.
DR InterPro; IPR002659; Glyco_trans_31.
DR Pfam; PF01762; Galactosyl_T.1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Signal-anchor; Golgi stack; Polymorphism.
FT DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 5 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 26 371 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 143 143 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 248 248 W -> R (IN STRAIN PGN2).
FT VARIANT 344 344 G -> W (IN STRAIN BLG2/MSF).
SO SEQUENCE 371 AA; 41235 MW; 323A7FA56B723B3 CRC64;

Query Match 14.1%; Score 298.5; DB 1; Length 371;
Best Local Similarity 28.4%; Pred. No. 2.3e-17;
Matches 83; Conservative 49; Mismatches 115; Indels 45; Gaps 9;

QY 130 LLIDQPKC---AKKPELLLAISLTPHFARRQAIRESAGQESNAGNQTVRVFLGQTP 186
DQ 56 LLISNHACGSGGPPPLLLIVCTAPEHLNORNAIRASMGAIREARFVOTLFLGK-P 114
QY 187 PEDNHPDLSMLKFESEKHQDILMWNRYDTFFNLSLKEVFLRWVSTSCPDTEFVKGD 246
DQ 115 RROGLADLS---ESAHRDILQASFODSYRNLTKLTSLGNVWVKYCPMARYILKTD 170
QY 247 DVFVTHHILNYL-----NSLSKTKAKD-----LFTGDVHNAGPH 282
DQ 171 DVVNVPELVSELVQSGPSEQQKGEAEETTAIEEHRGQAVPLLYLGRVHVRPT 230
QY 283 ROKKLKYYIPEVWYS---GLYPYVAGGGFLYSGHLALRLYHTDQVHLYPIDDVTGMC 339
DQ 231 RTPESRHVSEELWPNWGGPPPYASGTGYVLSISAVOLILKVASRAPLPLEDVFGVS 290
QY 340 LQKGLGYPE---KHGFTFTDIEKNKNICSVDMLV-HSRKPOEMIDIW 387
DQ 291 ARRGGLAPTHCVKLAGATHYPLDR-----CCYKGLLTSHKVPQWQMEAW 336

RESULT 9
B3G4_RAT
ID B3G4_RAT STANDARD: PRT: 371 AA.
AC O88178;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Beta-1,3-galactosyltransferase 4 (EC 2.4.1.62) (Beta-1,3-GalTase 4)
DE (Beta3Gal-T4) (B3Gal-T4) (Ganglioside galactosyltransferase) (UDP-
DE galactose:beta-N-acetyl-galactosamine-beta-1,3-galactosyltransferase)
DE (GAL-T2).
GN B3GALT4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL
RP STAGE.
RC TISSUE=Brain;
RX MEDLINE=97460055; PubMed=9312075;
RA Miyazaki H., Fukumoto S., Okada M., Hasegawa T., Furukawa K.,
RA Furukawa K.;
RT *Expression cloning of rat cDNA encoding UDP-galactose:GD2 beta1,3-
RT galactosyltransferase that determines the expression of
RT galib/GM1/GAL1.
RT GD1b/GM1/GAL1.
RL J. Biol. Chem. 272:24794-24799(1997).
CC -!- FUNCTION: Involved in GM1/GD1b/GAL ganglioside biosynthesis.
CC -!- CATALYTIC ACTIVITY: UDP-galactose + N-acetyl-D-galactosaminyl-(N-

```

```

CC acetylneuraminyl)-D-galactosyl-D-glucosyl-N-acylsphingosine = UDP
CC + D-galactosyl-N-acetyl-D-galactosaminyl-(N-acetylneuraminyl)-D-
CC galactosyl-D-glucosyl-N-acylsphingosine.
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.
CC -!- TISSUE SPECIFICITY: Highly expressed in thymus, spleen, kidney and
CC testis and, to a lesser extent, in brain and liver.
CC -!- DEVELOPMENTAL STAGE: In the embryonic brain, expression begins at
CC day 12 and continues until birth. Expression is maintained at low
CC levels in adult brain.
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AB003478; BAA32045.1; -.
CC InterPro; IPR002659; Glyco_trans_31.
CC Pfam; PF01762; Galactosyl_T.1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Signal-anchor; Golgi stack.
FT DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 5 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 26 371 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 143 143 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SEQUENCE 371 AA; 41254 MW; 327FB76FDACF131 CRC64;

Query Match 14.1%; Score 298.5; DB 1; Length 371;
Best Local Similarity 27.7%; Pred. No. 2.3e-17;
Matches 81; Conservative 51; Mismatches 115; Indels 45; Gaps 9;

QY 130 LLIDQPKC---AKKPELLLAISLTPHFARRQAIRESAGQESNAGNQTVRVFLGQTP 186
DQ 56 LLIPNPQAGSGGPPPLLLIVCTAPEHLNORNAIRASMGAIREARFVOTLFLGK-P 114
QY 187 PEDNHPDLSMLKFESEKHQDILMWNRYDTFFNLSLKEVFLRWVSTSCPDTEFVKGD 246
DQ 115 MQQGFADLAS---ESAAQGDVLQASFODSYRNLTKLTSLGNVWVKYCPMARYILKTD 170
QY 247 DVFVTHHILNYL-----NSLSKTKAKD-----LFTGDVHNAGPH 282
DQ 171 DVVNVPELVSELVQSGPSEQQKGEAEETTAIEEHRGQAVPLLYLGRVHVRPT 230
QY 283 ROKKLKYYIPEVWYS---GLYPYVAGGGFLYSGHLALRLYHTDQVHLYPIDDVTGMC 339
DQ 231 RTPESRHVSEELWPNWGGPPPYASGTGYVLSISAVOLILKVASRAPLPLEDVFGVS 290
QY 340 LQKGLGYPE---KHGFTFTDIEKNKNICSVDMLV-HSRKPOEMIDIW 387
DQ 291 ARRGGLAPTHCVKLAGATHYPLDR-----CCYKGLLTSHKVPQWQMEAW 336

RESULT 10
B3G4_HUMAN
ID B3G4_HUMAN STANDARD: PRT: 378 AA.
AC O96024;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Beta-1,3-galactosyltransferase 4 (EC 2.4.1.62) (Beta-1,3-GalTase 4)
DE (Beta3Gal-T4) (B3Gal-T4) (Ganglioside galactosyltransferase) (UDP-
DE galactose:beta-N-acetyl-galactosamine-beta-1,3-galactosyltransferase)
DE (GAL-T2) (GalT4).
GN B3GALT4 OR GALT4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```



```

DR EMBL: M63636; AAA63533.1; -.
DR PIR: A49750; A49750.
DR HSP: P00722; IBLG.
DR InterPro: IPR006101; Glyco_hydro_2.
DR InterPro: IPR006102; Glyco_hydro_2Ig.
DR InterPro: IPR006104; Glyco_hydro_2SB.
DR InterPro: IPR006103; Glyco_hydro_2T1M.
DR InterPro: IPR004200; Glyco_hydro_42C.
DR InterPro: IPR004199; Glyco_hydro_42N.
DR Pfam: PF02930; Bgal_small_C; 1.
DR Pfam: PF02929; Bgal_small_N; 1.
DR Pfam: PF00703; Glyco_hydro_2; 1.
DR Pfam: PF02836; Glyco_hydro_2_C; 1.
DR Pfam: PF02837; Glyco_hydro_2_N; 1.
DR PRINTS: PR00132; GLYDRLASE2.
DR PROSITE: PS00719; GLYCOSYL_HYDROL_F2.1; 1.
DR PROSITE: PS00608; GLYCOSYL_HYDROL_F2.2; 1.
DR Hydrolase: Glycosidase.
KW PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 458 458 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 546 546
SQ SEQUENCE 1026 AA; 117045 MW; B453AFAD68F08C98 CRC64;

Query Match
Best Local Similarity 5.2%; Score 110; DB 1; Length 1026;
Matches 76; Conservative 56; Mismatches 113; Indels 120; Gaps 20;

QY 63 QEKLNRQYNILSKLTNQTGEARLSNISHLNICEPDILRVTSV-----VIGFNN----- 111
DB 45 RQSLNGKWKIHYAONTNQV-----LKDIFYKTEPDETDLNFINVPGHLELQGGFSGPOYVNT 99
QY 112 -LPDRFDLRLCRNYSLLIDOPDKCAKPFLLAIAKSLTPHFAARQAIRESWGQESN 170
DB 100 QYWDGKEFL-----RPPQVPOES---NAVASYVHFNLDAIKDKKVFISF 143
QY 171 AGNOTVVRVFLGQ-----TPEDNHPDLSMLKFESEKHQDILMNNYRDIFNLS 221
DB 144 QGVATSFVWYNGFVCGYSEDSFTPSEF---E*SDYL-VEGDKNLAVAVRYSTA----- 194
QY 222 LKEVLFRLWYSTCPDTEFVKGDGDDVFVNTHTILN--YLSLSKTKAKDLFI-GDV--- 275
DB 195 -----SWL-----EDDFWRLYGIFRDVLYLAIPKVVODLFLVKGDYDYO 234
QY 276 -----IHNAGPHRDKKLYIPEWYSGLYGVPYAGGGGLYSGHLALRYHI----- 322
DB 235 TKAQQLDILKTVGDYEDKKIKYVLSL--YEGIVT--EGDASVNGDGLSVLENLKIFP 296
QY 323 --TDQVHLXPI-----DDVTGMCLOKGLVPEKHKGFRTFIEKNKNKNCISYVDLML 374
DB 291 WSAESPKLYDLILHVLDD-----QWVEVPVK-VGFRFRFEIKDK-----LML 332
QY 375 VHSRK 379
DB 333 LNGKR 337

RESULT 12
MDS3_YEAST
ID MDS3_YEAST STANDARD; PRT; 1487 AA.
AC P53094;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE MDS3 protein (MCK1 dosage suppressor 3).
GN MDS3 OR YGL197W OR G1307.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1679;
RX MEDLINE=97197971; PubMed=9046087;
RA Coglievina M., Klima R., Bertani I., Delneri D., Zaccaria P.,

```

```

RA Bruschi C.V.;
RT "Sequencing of a 40.5 kb fragment located on the left arm of
RI chromosome VII from Saccharomyces cerevisiae.";
RL Yeast 13:55-64(1997).
RN [2]
RP CHARACTERIZATION.
RA Li M.B., Nelgeborn L.;
RL Unpublished observations (XXX-1997).
CC !- FUNCTION: NOT KNOWN; NEGATIVE REGULATOR OF EARLY MEIOTIC GENE
CC EXPRESSION.
CC !- SIMILARITY: Contains 3 Kelch repeats.
CC !- SIMILARITY: TO YEAST YER132C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X91837; CAA62947.1; -.
DR EMBL: Z72719; CAA96909.1; -.
DR PIR: S62048; S62048.
DR TRANSFAC: T03484; -.
DR SGD: S0003165; MDS3.
DR GO: GO:0005737; C:cytoplasm; IDA.
DR GO: GO:0030435; P:sporulation; IGI.
DR InterPro: IPR006652; Kelch_rep.
DR Pfam: PF01344; Kelch; 2.
KW Kelch repeat; Repeat; Meiosis.
FT REPEAT 171 226 KELCH 1.
FT REPEAT 234 287 KELCH 2.
FT REPEAT 371 419 KELCH 3.
SQ SEQUENCE 1487 AA; 167073 MW; 768AEFBAB796E447 CRC64;

Query Match
Best Local Similarity 4.9%; Score 104; DB 1; Length 1487;
Matches 77; Conservative 59; Mismatches 120; Indels 158; Gaps 18;

QY 19 VFIYFIMEVSKSSQKNGKEVILPKFKWKISTPPEAYNNRQEKLNQYNPILSMLI 78
DB 172 LYIFGLTVSPQSGYE-----LIATNELWKLDLNKKW-----SLLS 208
QY 79 NOTGAGRLSNISHLNICEPDILRVTS--VWTFGNLPDRFKDFLLY----- 122
DB 209 DDPQIARRFNHTMHVKNENNDRDKLIIVGGLNNMDQPVKKIDIIYNISQNCWHPKTI 268
QY 123 -----LRCNYSLLIDOPDKCAKPFLLAIAKSLTPHFAARQAIRESWG 166
DB 269 QPMEITTVNGIPLALSQDNFISILVENNE--ANVPALAF-----YMRSDQIDEVLG 318
QY 167 QESN--AGNQTIVRVFLGQTP-----PEDNHPDLSMLKFESEKHQD 207
DB 319 KDSSKIKENSPIVALPLLSSESQIGRMPSNPALPKLLNVPEYELLAPTGDYVGF-----N 372
QY 208 ILMW-----NYRDTFENLSLKEVLFUR--VSTSCPDE-----FVEKG----- 244
DB 373 IIGGFHPNYQSSNFCFIYDINSKWSRVRTACDCKINHRFRWRFVWVWVSHQTILLG 432
QY 245 --DDDFVNTHTILNLYNLSL-----KTKAKDLFIGDVIHINAGPHRDKK 286
DB 433 TKTDYDYSVQVRFDLSTFGLPLVNFNFKTIQLPHHKISASSLPI--PIENFAKHKDT 490
QY 287 LK-----YYIPEVYVYSGL---YPPTAGGGGLYSGHLALRLY 320
DB 491 LKKVSFTSATSQFENYIRIYAPPLEMSIOSVFPFYA-----MVLGKDALEIY 539

RESULT 13
YAW6_SCHPO
ID YAW6_SCHPO STANDARD; PRT; 453 AA.
AC Q10181;

```

[illegible]

RESULT 15

```
YZ39_METJA
ID YZ39_METJA STANDARD: PRT: 351 AA.
AC Q60294;
DT 01-NOV-1997 (Rel. 35, Created)
DI 01-NOV-1997 (Rel. 35, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJEC139.
GN MJEC139.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L77118; AAC37108.1; -.
DR PIR: F64514; F64514.
DR TIGR: MJEC139; -.
DR Pfam: PF04473; DUF553; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 91 111 POTENTIAL.
SQ SEQUENCE 351 AA: 41437 MW; AACBF7C32BC213B0 CRC64;

Query Match
Best Local Similarity 4.5%; Score 95.5; DB 1; Length 351;
Matches 54; Conservative 27; Mismatches 35; Indels 95; Gaps 15;

QY 216 TFFNLS-----LKEYLF---LRWVSTSCPTEFVGDDDFVYTHHILNLSLS 263
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 TFFTLKCDKSLKLGKGLFTLSLWIST------LKTIL 156
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 264 KTK----AKDL--FGDVHNAQPHDKLKYIPEVYVSGLYPPYAGGGFLISCHLALR 318
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 157 KTKYAICRDYAKLTSAILHNL-----NIKHYF-----LVYPTHAVA 193
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 319 L-----YHITDQ-VHLYPIDDVTGMCLOKGLVPEKHGKGFRTFDIEE-----KN 362
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 194 VKIDYVYVIDOKLPIYKID-----VNLKKLG--KEKYIYTPVDIYNSKLKFEVKKYKN 246
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 363 KNNICSYV--DLMLVHSRKPQEMIDWSQLQ 391
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 247 ENNLKSEISDIL----RKIEE--DVKKELQ 271
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: October 20, 2003, 23:02:32
Job time : 14.5 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 22:56:05 ; Search time 18.5 Seconds
(without alignments)
2063.728 Million cell updates/sec

Title: US-09-804-357B-14
Perfect score: 2123
Sequence: 1 MSVGRRRKLLGILMMANVF.....RKQEMIDWSQSAHLKC 397
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	238	11.2	739	T10648	hypothetical prote
2	215	10.1	507	T18639	hypothetical prote
3	207	9.8	345	T20031	hypothetical prote
4	194	9.1	657	T20031	protein T7N9.18 [i
5	193.5	9.1	642	T20031	hypothetical prote
6	192.5	9.1	262	T20203	hypothetical prote
7	187	8.8	357	T20029	hypothetical prote
8	187	8.8	376	T20876	hypothetical prote
9	184	8.7	384	T20446	hypothetical prote
10	179	8.4	332	T20028	hypothetical prote
11	169.5	8.0	684	T86394	hypothetical prote
12	154.5	7.3	376	T24925	protein T24P13.20
13	153.5	7.2	399	T86453	hypothetical prote
14	145	6.8	395	T86458	avr9 homolog F9L1
15	140	6.6	398	T86353	probable elicitor
16	137.5	6.5	338	T24743	protein F2E2.6 [im
17	137.5	6.5	401	T86251	hypothetical prote
18	137	6.5	348	T31918	hypothetical prote
19	130	6.1	349	T24744	hypothetical prote
20	128.5	6.1	325	T24762	hypothetical prote
21	128.5	6.1	406	T204817	hypothetical prote
22	127.5	6.0	368	T15096	hypothetical prote
23	126.5	6.0	359	T87937	protein F14B6.6 [i
24	126.5	6.0	385	T20879	hypothetical prote
25	121	5.7	279	T20443	hypothetical prote
26	121	5.7	548	T86456	unknown protein [i
27	116.5	5.5	479	T86187	hypothetical prote
28	114.5	5.4	560	T86214	hypothetical prote
29	113.5	5.3	253	T22387	hypothetical prote

30	113.5	5.3	409	2	A84733	hypothetical prote
31	112	5.3	455	2	T29555	hypothetical prote
32	111.5	5.3	404	2	A86186	hypothetical prote
33	110	5.2	1026	2	A49750	beta-galactosidase
34	108	5.1	334	2	T32256	hypothetical prote
35	107.5	5.1	414	2	B96808	protein F28K19.2 [
36	107	5.0	491	2	S58330	probable membrane
37	105.5	5.0	723	2	H85092	hypothetical prote
38	105	4.9	353	2	C96573	protein F12M16.19
39	104	4.9	1487	2	S62048	probable membrane
40	102.5	4.8	453	2	T38707	probable initiator
41	99.5	4.7	544	2	G64483	hypothetical prote
42	99.5	4.7	1024	2	G72041	exodeoxyribonuclea
43	99.5	4.7	1024	2	D81624	exodeoxyribonuclea
44	99.5	4.7	1024	2	F86582	exodeoxyribonuclea
45	98.5	4.6	489	2	T13026	hypothetical prote

ALIGNMENTS

RESULT 1

T10648
hypothetical protein T13K14.220 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
C:Accession: T10648
R:Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Le
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z16991
A:Accession: T10648
A:Molecule type: DNA
A:Residues: 1-739 <BEV>
A:Cross-references: EMBL:AL080282; GSPDB:GN00062; ATSP:T13K14.220
A:Experimental source: cultivar Columbia; BAC clone T13K14
C:Genetics:
A:Gene: ATSP:T13K14.220
A:Map position: 4
A:Introns: 59/3; 86/1; 358/1; 438/3; 535/3; 578/3; 634/3; 670/3

Query Match 11.2%; Score 238; DB 2; Length 739;
Best Local Similarity 25.4%; Pred. No. 1.6e-11;
Matches 68; Conservative 56; Mismatches 108; Indels 36; Gaps 10;

Qy	142	PF-LLLAIKSLTPTFARRQAIRESWGQESNAGNOTVVRVFLLGQTTPEDNHPDLSMLKF	200
Db	492	PFRLFMGVLSATNHFSEMAVRKTMQHPSTKSDVVAREFFVALNP----RKEVNAMLKK	547
Qy	201	ESERHQDILMNYRDTFFNLSLKEVLEFLRWVSTSCPDTEFVKGDDDDVFVNTTHILNYLN	260
Db	548	EAIFYGDIVILPFMDRYELVVLKTIACEFGNVTAP---YIMKDDDTFIRVESILKQID	604
Qy	261	SLSTKAKDIFIGDVINAGPHRDKLKY---YIPEVYSGLYPPYAGGGFYLSGHLAL	317
Db	605	GVSPEKS--LYMGNLNLHRPLRTGKWTWTWEENPEAV----YPPYANGPGYIISNIA-	657
Qy	318	RLYHITDQ----VHLYPIDDVTGMCLOKGLVPEKHKGFTFDIEEKNKNICSY---	369
Db	658	--KIIVSQNSRHKLRLFKMEDVSMGLVQFNASMQ-----PVEYSHSWKFCQYGT	707
Qy	370	VDMLVSHSRKPQEMIDIWSQSAHLAK	397
Db	708	LNYYTAHYQSPSQMCLMDNLLKGRPQC	735

RESULT 2

T18639
hypothetical protein B0024.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18639
R:McMurray, A.
submitted to the EMBL Data Library, April 1996

A:Reference number: Z19001

A:Accession: T18639

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-507 <WIL>

A:Cross-references: EMBL:Z71178; PIDN:CAA94876.1; GSPDB:GN00023; CESP:B0024.3

A:Experimental source: clone B0024

C:Genetics:

A:Gene: CESP:B0024.3

A:Map position: 5

A:Introns: 69/2; 118/2; 143/2; 181/1; 246/2; 281/3; 393/2; 437/3

Query Match 10.1%; Score 215; DB 2: Length 507;

Best Local Similarity 22.6%; Pred. No. 7.6e-10;

Matches 91; Conservative 73; Mismatches 156; Indels 82; Gaps 16;

Qy 58 YNRE--QEKLNRYNPILS-----MTNOTGEAGR 86

Db 125 YDSRYAQEKISPTFFVMSHLPMDLDPDYFLDLDQCLIVVECTVVIILENTYSNVTV 184

Qy 87 LSNISHLNCEPDLRVTSVTVGNNLPDRFKD-----FLLYLRGRNYSL-LIDQPK-C 138

Db 185 FLNLDIRSYSDAPIFEVTSNFSKSHLSKOGSTNQFYHAQKDNHTYQFTVYKQOC 244

Qy 139 AKKPFLLAIAKSLTPHFARRQAIRESWGQESNA--GNQTVVRFLLGQTPEDNHPDLS 195

Db 245 SNTKLIQITLSPAGNFDIRQAIRETWANNPNSEHVANNDRVIRISLKSNE-----FLN 300

Qy 196 DMLKFESEKHODILMNYKXDTFNLSLKEVLEFLRVWVTSCTPTEVFVGDDSDVFNTHH; 255

Db 301 FALQKEIERFDDMIVDLYESYELLTKVHAISLYKQSHCOLADPQLKIDDDMAYVDMGL 360

Qy 256 LNYLSLSKTKAKDLFGIVGNAGPHROKKLYIPEVVIS--GLYPYVAGGGFLYSGH 314

Db 361 YRSLEDKKQASINGI--SGLIWNKSPVREKKHNVYPTLYSEKFFPPYIDPIVLIQKN 419

Qy 315 LAURLYHITDQVHLYPIDVY--TGMCLQKLGVLPEKHKGFRTFDIEKN--KNNICSYVDL 372

Db 420 AVPRMLEEAKNYNQWIEDVFVTGV---IG-----KALKIKQINWANHLLRYVIE 466

Qy 373 ML-----VHSRK-POEMIDINSQLSAHLKC 397

Db 467 LIPSRLKCSKGGVPLIYAVHNMKGPQNIHQYQKLGK--VKC 506

RESULT 3

T20031

hypothetical protein C47F8.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000

C:Accession: T20031; T22806

R:McLay, K.

submitted to the EMBL Data Library, December 1997

A:Reference number: Z19212

A:Accession: T20031

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-345 <WIL>

A:Cross-references: EMBL:AA09246; PIDN:CAAL5841.1; GSPDB:GN00019; CESP:C47F8.3

A:Experimental source: clone C47F8

R:Kershaw, J.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19617

A:Accession: T22806

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-345 <WIL>

A:Cross-references: EMBL:Z81553; PIDN:CAB04502.1; GSPDB:GN00019; CESP:C47F8.3

A:Experimental source: clone F56H6

C:Genetics:

A:Gene: CESP:C47F8.3

A:Map position: 1

A:Introns: 23/1; 49/1; 169/3; 236/2; 278/3; 308/3

C:Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

Query Match 9.8%; Score 207; DB 2: Length 345;

Best Local Similarity 23.1%; Pred. No. 2.1e-09;

Matches 86; Conservative 72; Mismatches 130; Indels 84; Gaps 16;

Qy 48 FWKISTPPEAYNREBEKLNRYNPILSMLTQNGAGRLSNISHLNCEPDLRVTSVVT 107

Db 24 FWKLQPNPKSSFYLEEHQVSGWN--ISTLSRPNIDFG-----SSFIV 64

Qy 108 GFNNLPDRKDFLLYL--RCRNYSLLDQPKCAKPKFLLAIAKSLTPHFARRQAIRESWG 166

Db 65 SPANIHKSOK--WYLPKMKNVYTKHD-----ILMLVVSXTKNFARNVLRSTWM 112

Qy 167 QESNAGNQTVVR---VFLGQTPPEDNHPDLSMLKFESEKHODILMNYRRTFNLSLK 223

Db 113 NKENSEMKSGRMHAFVGLVPGDN---LAKLVLEAEIHGDMVVDLEDYDNLDPFK 169

Qy 224 EVLFLRWVSTSCPDTEFVEKGGDDVFVNTHTIL---NYLNS--LS---KTKAKDLFI 272

Db 170 TLALLIGTSKASQFKIIGKIDDDVMFFDQDLPMLDRNFVNSNTLSYGHUSTAEELVL 229

Qy 273 GDVIHNAHPHROKKLYIPEVVIS--GLYPYVAGGGFLYSGHLA--LRLYHITDQVHLYP 330

Db 230 -----RNKTEPWYVPETAYNCTVYVYVMGPIYLVTKDAASILDONAHQOFMTV 279

Qy 331 IDDVYTCMLQKLGVLPEKHKGFRTFDIEB--KNNKNCISYVDLMLVH----- 376

Db 280 EDALITGITIAQKLGIGI-----RRYSLPNVFRHRNDITEGDDVLAHNVQTKNDSEYKSI 331

Qy 377 -SRKPOEMIDIW 387

Db 332 FTKKLSGTSFW 343

RESULT 4

G86397

protein T7N9.18 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: G86397

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Abo

ansen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, D.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marz

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tall

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: G86397

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-657 <STO>

A:Cross-references: GB:AE005172; NID:g8778858; PIDN:AAF79857.1; GSPDB:GN00141

C:Genetics:

A:Gene: T7N9.18

A:Map position: 1

Query Match 9.1%; Score 194; DB 2: Length 657;

Best Local Similarity 27.4%; Pred. No. 6e-08;

Matches 66; Conservative 43; Mismatches 106; Indels 26; Gaps 9;

Qy 158 ROATRESWGQESNAGNQTVVRVFLGQTPPEDNHPDLSMLKFESEKHODILMNYRRTF 217

Db 425 RMAVRKSMQKLVSRSSKVVARFEVALHARKEVNVND---LKKEAYFGDIVIVPMYDH 480

Qy 218 FNLSLKEVLEFLRW-VSTSCPDTEFVEKGGDDVFVNTHTILNLSLSTKAKD-LFIGDV 275

Db 481 DLVVLKTAICEYGVNTVA--AKYVMKCDODTFRVDAV---IOEAKYKGRSLYIGNI 535

Db 157 GDELKRLTEKGIIMRFVIGHS-----SSPGVLDHTIEAEEQHKDFRLNHHIEGYHLS 212
QY 222 LKEVLFL-----RWVSTSCPOTEFVFKGDDVFNTHILNLS-LSKTKAKD-LFQGD 274
Db 213 SKTOIFYSAVAKW-----DADFVIKVDVDDVHN-----LGMGSTLARHRSKPRVIGC 262
QY 275 VTHNAGP-HRDKKLYIPEVYVS-----LYPPYAGGGFLYSGHLALRLYHITDQVHL 329
Db 263 M--KSGPVLQAGVYHEPEYKFGEGNKYFRHATGQIYAIKSKDLATYISVNRQLLHKY 320
QY 330 PIDDVVTGMLQKGLVPEKHKGFRT-----FDTEENK-----NNICSYV 370
Db 321 ANEDVSLGSNF--IGLDVE-HIDDRSLCCGPLDCEWKGAGNCAASFDSGCSICKSV 377
QY 371 DMLV-VHSRRKPOEMDIW 387
Db 378 DRMLEVHORCGEGGAIW 395

RESULT 14
A86458
probable elicitor response protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 27-Nov-2001
C:Accession: A86458
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luross, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.W.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86458; MUID:21016719; PMID:11130712
A:Accession: A86458
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-395 <STO>
A:Cross-references: GB:AE005172; NID:g10645491; PIDN:AAG21603.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 6.8%; Score 145; DB 2: Length 395;
Best Local Similarity 27.0%; Pred. No. 0.00034;
Matches 74; Conservative 43; Mismatches 99; Indels 58; Gaps 16;

QY 114 DRPKDFLLYLCRNYSLLIDQDKAKKDFLLLAISLTPHPARRQAIRESW---GQ--- 167
Db 104 DRSEFWSESAKNSQRL-----QKFAVIGINTAFSSKKRDSVRQTMPTGEKLLK 155
QY 168 --ESNAGNQTVVVFLLGOTPPEDNHPDLSDMLKFESEKHODILMNYRDTFFNLSLKEV 225
Db 156 KIEKEG--IVVR-FVIGHSATPGG--VLDKAIDEDSEKDFLKLKIEGYHQLSTKR 210
QY 226 LFLRWSTSCPTEFFVFKGDDVFNTHILNLSLTKAK-----DLFTGVDVHNAG 280
Db 211 LYFS-TATAMYDAEFYKVDVDDVHN-----LGMVLTTLARYQSRPRYIGCM--KSG 260
QY 281 PHRDKK-LKYIPEVYVS-----LYPPYAGGGFLYSGHLALRLYHITDO--VHLYPTDD 333
Db 261 PVLSQKGVYHEPEYKFGEGNKYFRHATGQIYAIKSKDLA--TYISTNQGLHRYANED 318
QY 334 VYTGMLQKGLVPEKHKGFRTDIEKNKNIC 367
Db 319 VSLGANW-LGL-----EVEHVDERSMC 339

RESULT 15
B86353
protein F2E2.6 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B86353
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alo,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 616-620, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luross, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tall,
ker, M.; Wu, D.; Yu, G.; Fraser, C.W.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: B86353
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-398 <STO>
A:Cross-references: GB:AE005172; NID:g9280694; PIDN:AAF86563.1; GSPDB:GN00141
C:Genetics:
A:Gene: F2E2.6
A:Map position: 1

Query Match 6.6%; Score 140; DB 2: Length 398;
Best Local Similarity 25.1%; Pred. No. 0.00089;
Matches 62; Conservative 40; Mismatches 95; Indels 50; Gaps 11;

QY 141 KPFLLLAIKSLTPHPARRQAIRESW---GQ---ESNAGNQTVVVFLLGOTPPEDNHPDLS 195
Db 125 KVFVIGINTAFSSRRKRDSRLRETWMPOGEKLEKLEKGIIVKEMIGHSSITPNSM--LD 182
QY 196 DMLKFESEKHQDILMNYRDTFFNLSLKEVFL-----RWVSTSCPTEFFVFKGDDVVF 250
Db 183 KEIDSEDAQYNDFRLDHVEGYINLSAKTSFESSAVAKW-----DAEFYVKIDDDVHV 236
QY 251 NTHIILNLSLTKAKD-----LFIGDVIHNAGPHRDKK-LKYIPEVYVS-----LY 300
Db 237 N-----LGTLASTLASHRSKPRVYIGCM--KSGPVLTKTKAKYREPEFVKFGEENKY 287
QY 301 PPVAGGGFLYSGHLALRLYHITDOVHLYPIDDVYTGMCLOKGLVPEKHKGFRTDIEE 360
Db 288 FRHATGQIYAIKSKDLATYISNQPILHRYANEDVTLG-----SWFICLEVEQ 334
QY 361 KNKNIC 367
Db 335 IDDRNFC 341

Search completed: October 20, 2003, 23:04:47
Job time : 19.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 22:55:10 ; Search time 42 Seconds

(without alignments)
2439.212 Million cell updates/sec

Title: US-09-804-357b-14

Perfect score: 2123

Sequence: 1 MSVRRRRRIKILGILMANVF.....RKPEMIDISQLQSAHLKC 397

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_plant:*
- 10: sp_protist:*
- 11: sp_rhodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacterioph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1872	88.2	397	11 Q91V18	Q91V18 mus musculus
2	1866	87.9	397	11 Q92222	Q92222 mus musculus
3	1232.5	58.1	406	13 Q8UWM0	Q8UWM0 brachydanio
4	1145.5	54.0	412	13 Q8UWM3	Q8UWM3 brachydanio
5	680.5	32.1	412	13 Q8UWM2	Q8UWM2 brachydanio
6	668	31.5	401	4 Q8NFLO	Q8NFLO homo sapien
7	660.5	31.1	397	11 Q8K437	Q8K437 mus musculus
8	651.5	30.7	397	11 Q8K0J2	Q8K0J2 mus musculus
9	651	30.7	418	13 Q8UWM4	Q8UWM4 brachydanio
10	635.5	29.9	350	11 Q923H4	Q923H4 mus musculus
11	632.5	29.8	353	4 Q8N934	Q8N934 mus sapien
12	628.5	29.6	353	4 Q96QH5	Q96QH5 homo sapien
13	628.5	29.6	353	4 Q8WY02	Q8WY02 homo sapien
14	628.5	29.6	353	4 Q8WWR5	Q8WWR5 homo sapien
15	628.5	29.6	378	4 Q8C0J1	Q8C0J1 homo sapien
16	628.5	29.6	378	4 Q8N5W4	Q8N5W4 homo sapien

17	628.5	29.6	388	4 Q8ND21	Q8ND21 homo sapien
18	616.5	29.0	384	4 Q8TDX1	Q8TDX1 homo sapien
19	615.5	29.0	390	13 Q8UWM1	Q8UWM1 brachydanio
20	603.5	28.4	377	4 Q96EK0	Q96EK0 homo sapien
21	600	28.3	372	4 Q9C0J2	Q9C0J2 homo sapien
22	573	27.0	370	11 Q8BK98	Q8BK98 mus musculus
23	573	27.0	399	11 Q8V116	Q8V116 mus musculus
24	572.5	27.0	372	11 Q8R0U2	Q8R0U2 mus musculus
25	569.5	26.8	372	11 Q9D722	Q9D722 mus musculus
26	549.5	25.9	277	4 Q8TAZ4	Q8TAZ4 homo sapien
27	549	25.9	389	11 Q8R319	Q8R319 mus musculus
28	507	23.9	374	4 Q8WWR6	Q8WWR6 homo sapien
29	479.5	22.6	326	4 Q9Y526	Q9Y526 homo sapien
30	479.5	22.6	326	6 Q9MYM7	Q9MYM7 pongo pygma
31	479.5	22.6	326	11 Q54904	Q54904 mus musculus
32	462	21.8	305	11 Q91V52	Q91V52 mus musculus
33	462	21.8	305	11 Q920V5	Q920V5 mus spicille
34	430.5	20.3	409	11 Q920V2	Q920V2 mus spicille
35	427.5	20.1	422	4 Q43825	Q43825 homo sapien
36	427	20.1	409	11 Q91V19	Q91V19 mus musculus
37	427	20.1	409	11 Q920V3	Q920V3 mus musculus
38	427	20.1	409	11 Q91V58	Q91V58 mus musculus
39	427	20.1	409	11 Q920V4	Q920V4 mus musculus
40	427	20.1	409	11 Q91VE9	Q91VE9 mus musculus
41	427	20.1	422	11 Q54905	Q54905 mus musculus
42	427	20.1	422	11 Q8CBX4	Q8CBX4 mus musculus
43	427	20.1	422	11 Q8BH19	Q8BH19 mus musculus
44	425.5	20.0	422	4 Q9BZQ9	Q9BZQ9 homo sapien
45	416.5	19.6	378	4 Q9BYG0	Q9BYG0 homo sapien

ALIGNMENTS

RESULT 1

Q91V18					
ID	Q91V18	PRELIMINARY;	PRT;	397	AA.
AC	Q91V18				
DT	01-DEC-2001 (Tremblrel. 19, Created)				
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)				
DT	01-OCT-2002 (Tremblrel. 22, Last annotation update)				
DE	Beta-1.3-N-acetylglucosaminyltransferase (Beta-1.3-N-				
DE	acetylglucosaminyltransferase 1).				
GN	B3GNT1				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C.B17;				
RA	Egan S.E., Cohen B.L., Sarkar M., Ying Y., Cohen S., Singh N.,				
RA	Wang W., Flock G., Goh T., Schachter H.,				
RT	"Molecular cloning and expression analysis of a mouse UDP-				
RT	GlcNAc:Gal(beta1-4)Glc(NAC)-R beta1.3-N-acetylglucosaminyltransferase				
RI	homologous to Drosophila melanogaster Brainiac and the beta1,3-				
RI	galactosyltransferase family.";				
RL	Glycoconj. J. 17:865-872(2000).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C.B17;				
RA	Egan S.E., Cohen B.L., Sarkar M., Ying Y., Cohen S., Singh N.,				
RA	Wang W., Flock G., Goh T., Schachter H.,				
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	T1SSOE-Breast tumor;				
RA	Strausberg R.;				
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AY043479; AAK95359.1; "				
DR	EMBL; BC009075; AAH09075.1; "				
DR	MGI; MGI:1889505; B3gnrl.				
DR	InterPro; IPR002659; Glyco_trans_31.				

```

DR Pfam: PF01762; Galactosyl_T; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 397 AA; 45883 MW; D8BBEA1866C1D106 CRC64;

Query Match 88.2%; Score 1872; DB 11; Length 397;
Best Local Similarity 87.4%; Pred. No. 2.2e-154;
Matches 347; Conservative 19; Mismatches 31; Indels 0; Gaps 0;

Qy 1 MSVGRRRKILGLILMANVFIYFIMEVSKSSQKNGKEVIIPKEFKWIKSTPPAYWN 60
Db 1 MSVGRRRKILGLILMANVFIYFIMEVSKSSQKNGKEVIIPKEFKWIKSTPPAYWN 60
Qy 61 REEKLNRQYNPILSLTNOTGEAGRLSNISHLNYCEPDLRVTSVVTGNNLPDRFKDFL 120
Db 61 REEKLNRQYNPILSLTNOTGEAGRLSNISHLNYCEPDLRVTSVVTGNNLPDRFKDFL 120
Qy 121 LYLRCRNYSLIDOPDKCAKPPFLLLAISKSLTPHFARRQAIRESWGQESNAGNQTIVRVF 180
Db 121 LYLRCRNYSLIDOPDKCAKPPFLLLAISKSLTPHFARRQAIRESWGQESNAGNQTIVRVF 180
Qy 181 LLGOTPPEDNHPDLSMDLKFSEKHQDILMNNYRTDFNLSLKEVLFRLRWVTSICPDTEF 240
Db 181 LLGOTPPEDNHPDLSMDLKFSEKHQDILMNNYRTDFNLSLKEVLFRLRWVTSICPDTEF 240
Qy 241 VFKGDDVFNVTNTHILNLSKRAKOLFIDGVIHNAQPHRDKKLYIPEVYSGLY 300
Db 241 VFKGDDVFNVTNTHILNLSKRAKOLFIDGVIHNAQPHRDKKLYIPEVYSGLY 300
Qy 301 PPYAGGGGFLYSGHLALRLYHITDQVHLYPIDDDVYTGMCLOKLGVLPEKHKGRTFDIEE 360
Db 301 PPYAGGGGFLYSGHLALRLYHITDQVHLYPIDDDVYTGMCLOKLGVLPEKHKGRTFDIEE 360
Qy 361 KNNKNCISYVDLMLVHSRKPQEMIDWSQLOSPNLKC 397
Db 361 KNNKNCISYVDLMLVHSRKPQEMIDWSQLOSPNLKC 397

RESULT 2
Q92222 ID O92222 PRELIMINARY; PRT; 397 AA.
AC O92222;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE N-acetylglucosaminide beta-1,3-N-acetylglucosaminyltransferase
DE (EC 2.4.1.149) (Poly-N-acetylglucosamine extension enzyme) (N-
DE acetylglucosaminyltransferase) (Beta3GNT).
DE B3GNT1 OR BETA3GNT.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=ICR; TISSUE=Neonatal Brain;
RX MEDLINE=99110903; PubMed=9892646;
RA Zhou D., Dinter A., Gutierrez Gallego R., Kamerling J.P.,
RA Villegenthart J.F.G., Berger E.G., Hennet T.;
RT "A beta-1,3-N-acetylglucosaminyltransferase with poly-N-
RT acetylglucosamine synthase activity is structurally related to beta-
RT 1,3-galactosyltransferases."
RL Proc. Natl. Acad. Sci. U.S.A. 96:406-411(1999).
RN [2]
RP REVISIONS.
RA Zhou D., Berger E.G., Hennet T.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CAN INITIATE THE SYNTHESIS OR THE ELONGATION OF THE
CC LINEAR POLY-N-ACETYLGLUCOSAMINOGLYCANS.
CC -!- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-
CC GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R = UDP + N-ACETYL-BETA-D-
CC GLUCOSAMINYL-1,3-BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R.
CC -!- COFACTOR: MANGANESE.
CC -!- PATHWAY: GLYCOSYLATION.

```

```

CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, LUNG, KIDNEY AND
CC TESTIS AND TO A LESSER EXTENT IN LIVER AND SKELETAL MUSCLE. NO
CC EXPRESSION IN SPLEEN.
DR EMBL: AF092050; AAD09763.2; -.
DR MGD: MGI:1889505; B3gnl1.
DR InterPro: IPR004659; Galactosyl_T; 1.
DR Pfam: PF01762; Galactosyl_T; 1.
KW Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW Glycoprotein; Golgi stack.
FT DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 8 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 29 325 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 30 30 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 397 AA; 45813 MW; 29BBEADF7B770A6B CRC64;
Query Match 87.9%; Score 1866; DB 11; Length 397;
Best Local Similarity 87.2%; Pred. No. 7.2e-154;
Matches 346; Conservative 19; Mismatches 32; Indels 0; Gaps 0;

Qy 1 MSVGRRRKILGLILMANVFIYFIMEVSKSSQKNGKEVIIPKEFKWIKSTPPAYWN 60
Db 1 MSVGRRRKILGLILMANVFIYFIMEVSKSSQKNGKEVIIPKEFKWIKSTPPAYWN 60
Qy 61 REEKLNRQYNPILSLTNOTGEAGRLSNISHLNYCEPDLRVTSVVTGNNLPDRFKDFL 120
Db 61 REEKLNRQYNPILSLTNOTGEAGRLSNISHLNYCEPDLRVTSVVTGNNLPDRFKDFL 120
Qy 121 LYLRCRNYSLIDOPDKCAKPPFLLLAISKSLTPHFARRQAIRESWGQESNAGNQTIVRVF 180
Db 121 LYLRCRNYSLIDOPDKCAKPPFLLLAISKSLTPHFARRQAIRESWGQESNAGNQTIVRVF 180
Qy 181 LLGOTPPEDNHPDLSMDLKFSEKHQDILMNNYRTDFNLSLKEVLFRLRWVTSICPDTEF 240
Db 181 LLGOTPPEDNHPDLSMDLKFSEKHQDILMNNYRTDFNLSLKEVLFRLRWVTSICPDTEF 240
Qy 241 VFKGDDVFNVTNTHILNLSKRAKOLFIDGVIHNAQPHRDKKLYIPEVYSGLY 300
Db 241 VFKGDDVFNVTNTHILNLSKRAKOLFIDGVIHNAQPHRDKKLYIPEVYSGLY 300
Qy 301 PPYAGGGGFLYSGHLALRLYHITDQVHLYPIDDDVYTGMCLOKLGVLPEKHKGRTFDIEE 360
Db 301 PPYAGGGGFLYSGHLALRLYHITDQVHLYPIDDDVYTGMCLOKLGVLPEKHKGRTFDIEE 360
Qy 361 KNNKNCISYVDLMLVHSRKPQEMIDWSQLOSPNLKC 397
Db 361 KNNKNCISYVDLMLVHSRKPQEMIDWSQLOSPNLKC 397

RESULT 3
Q9UWMO ID Q8UWMO PRELIMINARY; PRT; 406 AA.
AC Q8UWMO;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Beta-3-galactosyltransferase.
DE SSP5.
GN Brachydanio rerio (Zebrafish) (Danio rerio).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Chou C.-M., Leu J.-H., Huang C.-J.;
RA "A family of novel genes encoding beta-3-galactosyltransferase from
RT zebrafish, zssp5 mRNA."

```

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF321831; AAL32299.1; -

DR InterPro; IPR002659; Glyco_trans_31.

DR Pfam; PF01762; Galactosyl_T; 1.

KW Transferase; Glycosyltransferase.

SQ SEQUENCE 406 AA: 46953 MW: F928471BBDF0F1681 CRC64;

Query Match 58.1%; Score 1232.5; DB 13; Length 406;

Best Local Similarity 58.2%; Pred. No. 8.9e-99;

Matches 226; Conservative 70; Mismatches 87; Indels 5; Gaps 3;

QY 5 RRRKLLGILMMANVFYIETMEYSSKSSOENKCKGEV:IPKEFKWIKISTPPEAYWNRQE 64

DB 5 RRKVKVA--WMTWVFLFIVYVSRNAGKSSKNSLVLPKRFWAKDLPSPDAYNRQQ 62

QY 65 KLNROYNPILSMITNOIGE--AGRLSNISHINCEPDRLVTSVVTFGNLPDRFKDFLY 122

DB 63 QINYINNRLEKL-NYTDNLDPWLDNVSLSDCDPDYRTVTQVDYNSLPDRFKDFLY 121

QY 123 LRCNYSLLIDOPDKCAKPFLLAIKSLTPHFARROAIRSQESQESNAGNQTVVRFLL 192

DB 122 MRCRSYPIVDVQPNICKKQPFLLAIKSLVPHFDRROAIRSQESQESNAGNQTVVRFLL 181

QY 183 GQTPEDNHPDLSMLAFSEKSHODILMNNYRTFFNLSLKEVLFRLRWSTSCPDTERVF 242

DB 182 GNAATEDHFDPLSKMLHSESHIRDILOWDYRTFFNLTKEVLFLEWLSRCPGANFIF 241

QY 243 KGDDOVFNTHILNLSKTKAKDLFTGDIIVHAGPHRDKKLYIPEVYVSGLYPP 302

DB 242 KGDDOVFNTHIIFDNLNLSKAKARELFVGDVITNAGPHRDKKLYIPEVYVSGLYPP 301

QY 303 YAGGGGLYSGLALRLHYHTDOVHLYPIDVYTGMCLOKGLVPEKHKGFRFTDIEKN 362

DB 302 YAGGGGLYSGLAQLRLHNSKLVPLPIDVYTGMCMLWGLAPEKHKGFRFTDIEKY 361

QY 363 KNNITCSYVDMLVHSRKPQEMIDWISQL 390

DB 362 RDNACAYKSLMLVHPSQPMIKIWAWL 389

RESULT 4

Q8UWM3

ID Q8UWM3 PRELIMINARY; PRT: 412 AA.

AC Q8UWM3;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Beta-3-galactosyltransferase.

GN SSP2.

OS Brachydanio rerio (zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Danio.

OX NCBI_TaxID=7955;

RN [1]

RP SEQUENCE FROM N.A.

RA Chou C.-M., Leu J.-H., Huang C.-J.;

RT "A family of novel genes encoding beta-3-galactosyltransferase from

zebrafish, zssp2 mRNA";

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF321828; AAL32296.1; -

DR InterPro; IPR002659; Glyco_trans_31.

DR Pfam; PF01762; Galactosyl_T; 1.

KW Transferase; Glycosyltransferase.

SQ SEQUENCE 412 AA: 47440 MW: 14F24C26570B5F45 CRC64;

Query Match

Best Local Similarity 54.0%; Score 1145.5; DB 13; Length 412;

Matches 224; Conservative 58; Mismatches 111; Indels 7; Gaps 5;

QY 1 MSVGRRRRLKLGILMMANVFYIETMEYSSKSSQEKNG-KGEVIIPKEFKWIKISTPPEAYW 59

DB 1 MSNNWNSTKLLGVLLANFLIYITAVEVSRSHNDRNAFKNNKIPKE-FWKKTKVSTAFAW 59

QY 60 NREQEKLRQYNPIILSMITNOTGEAGRLSNISH-LNY-CEPDLRVTSVVTGFNNLPDRFK 117

DB 60 NREQERLDDIY---YLPVNASSELPHPFRGIPRWLNHSHCKHDVNVNTAIEDFNSLPDRFQ 116

QY 118 DFLLYLRCRNYSLLIDOPDKCAKPFLLAIKSLTPHFARROAIRSQESQESNAGNQTVV 177

DB 117 DFLLYMGCRSYPILITRAPKVCSPKPPYLLAIKSLAPHFDRROAIRSQESGRAGILDGRIA 176

QY 178 RVFLLGQTPEDNHPDLSMLKFESEKHODILMNNYRTFFNLSLKEVLFRLRWSTSCPD 237

DB 177 TVFLLGNTASTDHFDPDLSNWKHEALYGDVLOWDYRTFFNLTKEVLFLEWLSGSHCS 236

QY 238 TEFVEFKGDDVFNTHILNLSKTKAKDLFTGDIIVHAGPHRDKKLYIPEVYV 297

DB 237 AQYVEFKGDDVFNTHILNLSKTKAKDLFTGDIIVHAGPHRDKKLYIPEVYV 296

QY 298 GLYPYAGGGGLYSGLALRLHYHTDOVHLYPIDVYTGMCLOKGLVPEKHKGFRFTD 357

DB 297 GGYPPYAGGGGLYSGLALRLHYHTDOVHLYPIDVYTGMCLOKGLVPEKHKGFRFTD 356

QY 358 IBEKNKNNITCSYVDMLVHSRKPQEMIDWISQLOSAHLK 397

DB 357 IBEKHRENPCAYKGLILVHPSQPMIKIWSWINDPNAIC 396

RESULT 5

Q8UWM2

ID Q8UWM2 PRELIMINARY; PRT: 412 AA.

AC Q8UWM2;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Beta-3-galactosyltransferase.

GN SSP3.

OS Brachydanio rerio (zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Danio.

OX NCBI_TaxID=7955;

RN [1]

RP SEQUENCE FROM N.A.

RA Chou C.-M., Leu J.-H., Huang C.-J.;

RT "A family of novel genes encoding beta-3-galactosyltransferase from

zebrafish, zssp3 mRNA";

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF321829; AAL32297.1; -

DR InterPro; IPR002659; Glyco_trans_31.

DR Pfam; PF01762; Galactosyl_T; 1.

KW Transferase; Glycosyltransferase.

SQ SEQUENCE 412 AA: 47812 MW: 80C554675A2CB00A CRC64;

Query Match

Best Local Similarity 32.1%; Score 680.5; DB 13; Length 412;

Matches 123; Conservative 66; Mismatches 100; Indels 5; Gaps 2;

QY 109 FNNLPDRKDFLLYLRNYSLLIDOPDKCAKPFLLAIKSLTPHFARROAIRSQESQ 168

DB 113 FRFLSPRHFVLRHRCRYFPMNLNHPKCGGGVDLVVVKSVTEEDHRRRAVRKTWKE 172

QY 169 SNAGNQTVVRFLLGQTPEDNHPDLSMLKFESEKHODILMNNYRTFFNLSLKEVLF 228

DB 173 QEIOGLKIKTFLFLLGTAPAGKDSRLQALQVYEDRTYGDILQWDFMDTFFNLTKVNF 232

QY 229 RWVSTSCPDTEVFVKGDDVFNTHILNLSKTKAKDLFTGDIIVHAGPHRDKK 287

DB 233 RWFSLYCPDVPFIFKGGDDVFNTHILNLSKTKAKDLFTGDIIVHAGPHRDKK 292

QY 288 KYIYIPEVYVSGLYAGGGGLYSGLALRLHYHTDOVHLYPIDVYTGMCLOKGLV 347

DB 293 KYFIFRELKYDRYPYLGCGGFLMSSQVARKVFTVSEVELYPIDDVFGVCLQKLV 352

QY 348 EXHKGFRFTDIEKNKNNI---CSYVDMLVHSRKPQEMIDWISQLOSAHLK 397

[illegible]


```

Db 97 FQSLPFRFLAYRCHRVFPMNLNHPKCAQGVYMLVYVKSVITQCHDREVIQRTWGHE 156
Qy 169 ---SNAGNQTVVRVFLGQTPTPDNDNHPDLSMDLKFESEKHQDILMKNYKDTFFNLSLKEV 225
Db 157 WESAGLGRGAVRFLFLGTASKQERTHQOOLLAYEDRLYADILQWDFLDSFNLTLKEI 216
Qy 226 LFLRWVSTSPOTFEVFKGGDDVFNTHHILNLSLTKAKDPLFGDVIHNAHPHROK 285
Db 217 HFLKWLDIYCPNVPFVFKGGDDVFNTHHILNLSLTKAKDPLFGDVIHNAHPHROK 274
Qy 286 KLXYIPEVYV-SGLYPPYAGGGFLYSGHLALRLYHITDQVHLYPIDDVYGMCLQKLG 344
Db 275 DNYIYIPAVNYGATVPYAGGGFLYSGHLALRLYHITDQVHLYPIDDVYGMCLQKLG 334
Qy 345 LVPEKHKGFTFDIEEKNKNNI----CSYVDLMLVHSRKPOEMIDVWSLOSAHLKC 397
Db 335 VKPTGHEGFTFGISRVSRSSRMNKEPCFYRAMLVVHKLLPAELLAMMDLVHS-NLTC 390

RESULT 9
Q8UWM4 PRELIMINARY: PRT: 418 AA.
AC Q8UWM4;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Beta-3-galactosyltransferase.
GN SSP1.
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Chou C.-M., Liu J.-H., Huang C.-J.;
RT "A family of novel genes encoding beta-3-galactosyltransferase from
RT zebrafish, zsspl mRNA";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF321827; AAL32295.1; -.
DR InterPro; IPR002659; Glyco_trans_31.
DR Pfam; PF01762; Galactosyl_T; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 418 AA; 48936 MW; C7225B6259E1622B CRC64;

Query Match 30.78; Score 651; DB 13; Length 418;
Best Local Similarity 43.18; Pred. No. 3.3e-48;
Matches 127; Conservative 61; Mismatches 97; Indels 10; Gaps 5;

Qy 109 FNNLPDRFKDFLLYLCRCNYSLLIDOPDKCAKPFLLLAIAKSLTPHFARRQAIRESWGQE 168
Db 120 FTGPDNFKQFLYRHCYRFPMLNHPKCSGEIDLLIVIKSVITQDREVIQRTWGHE 179
Qy 169 SNAGNQTVVRVFLGQTPTPDNDNHPDLSMDLKFESEKHQDILMKNYKDTFFNLSLKEVIF 228
Db 180 QVNLGKRIKTLFLLGKSNLEERANHQLEEDYIYGDVTLQWDFMDSFNLTLKEHFL 239
Qy 229 RNVSTSPOTFEVFKGGDDVFNTHHILNLSLTKAKDPLFGDVIHNAHPHROK 288
Db 240 KWFSSYCPKQIYIPKGGDDVFNTHHILNLSLTKAKDPLFGDVIHNAHPHROK 297
Qy 289 YIYIPEVYV-SGLYPPYAGGGFLYSGHLALRLYHITDQVHLYPIDDVYGMCLQKLG 347
Db 298 YIYIPEVYV-SGLYPPYAGGGFLYSGHLALRLYHITDQVHLYPIDDVYGMCLQKLG 357
Qy 348 EKHKGFTFDIEEKNKNNI----ICSYVDLMLVHSRKPOEMIDVWSLOSAHLKC 397
Db 358 IKHNAFTFGL-VKNKTSRLNRPFCFFKSLIVVHKLLPDLMSMW-KLVNSDLIC 410

```

RESULT 10
Q923H4

```

ID Q923H4 PRELIMINARY: PRT: 350 AA.
AC Q923H4;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Bctal;3 N-acetylglucosaminyltransferase-4.
GN B3GNT4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhou D., Hennes T.;
RT "Mouse beta1.3 N-acetylglucosaminyltransferase-4.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY037786; AAK68856.1; -.
DR InterPro; IPR002659; Glyco_trans_31.
DR Pfam; PF01762; Galactosyl_T; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 350 AA; 39887 MW; 55A12083026D29B1 CRC64;

Query Match 29.9%; Score 635.5; DB 11; Length 350;
Best Local Similarity 41.58; Pred. No. 5.9e-47;
Matches 130; Conservative 59; Mismatches 105; Indels 19; Gaps 8;

Qy 91 SHLNVCPEPLRVSTVYVTFNLPDRFKDFLLYLCRCNYSLLIDOPDKCAKPFLLLAIAKS 150
Db 45 SHHSQSPNL--TVNTSLPSLRHRLFLYRCHRNFSILL-EPSECAKDTFLLLVKS 100
Qy 151 LTPHFARRQAIRESWGQESN-AGNQTVVRVFL--GQTPEDNHPDLSMDLKFESEKHQ 206
Db 101 QPAHIEQSAIRSTWGRAGSARGQLKLVFLGAGVPPV-----AQLLVYESWQDF 153
Qy 207 DILMKNYRDTFFNLKLEVLFLRWVYSTSCPDTFVFKGGDDVFNTHHILNLSLTK 266
Db 154 DILQWDFMDSFNLTLKEHVLQVMTAACTQAHFLKGGDDVFIHVPNVLEFLEGWD--P 211
Qy 267 AKDLFGDVIHNAHPHROKLLKYIPEVYVYSGLYPPYAGGGFLYSGHLALRLYHITDQ 325
Db 212 AODFLVGDVIRLARENRTKVFYFPMYRARIYPPYAGGGYVMSQATVRRHLHRAEMEE 271
Qy 326 VHLYPIDDVYGMCLQKLGVLPERKHGFRTEIDIEK-NKNNICSVYDMLVHSRKPOEMI 384
Db 272 AELFPIDDVYGMCLRLKLVTPIHAGFKTFGICQPLNRPDPCLYKGLLVHRLSPLEMW 331
Qy 385 DIWSLOSAHLKC 397
Db 332 TWWALVTDERLKC 344

RESULT 11
Q8N934 PRELIMINARY: PRT: 353 AA.
AC Q8N934;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DE 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ38427.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Brain;
RA Suzuki O., Sasaki N., Aotsuka S., Shoji T., Ichihara T., Shiohata N.,
RA Matsumoto K., Hirano M., Sano S., Nomura R., Yoshikawa Y.,
RA Matsumura Y., Moriya S., Chiba E., Momiyama H., Onogawa S.,
RA Kaeriyama S., Satoh N., Matsunawa H., Takahashi E., Kataoka R.,
RA Kuga N., Kuroda A., Satoh I., Kamata K., Takami S., Terashima Y.,
RA Watanabe M., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A.,

```



```
Qy 331 IDVVYTGMCLOKGLVPEKHGKFTFDIEEK-NKNNNICSYVDMLVHRSKPKQEMIDIWSQ 363
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 278 IDOVFGMCLRLGLSPMHAGKFTGIRRLDPLDPCLYRGILLVHRLSPLEMTWML 337
Qy 390 LQSAHLKC 397
:|||||
Db 338 VTDEGLKC 345

RESULT 14
ID Q8WRS PRELIMINARY; PRT; 353 AA.
AC Q8WRS;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Beta 1,3 galactosyltransferase.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bennett E.P.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DB EMBL; AJ278962; CAC82375.1; -.
DR InterPro; IPR002659; Glyco_trans_31.
DR Pfam; PF01762; Galactosyl_T; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 353 AA; 39863 MW; EAA7A35019C36D64 CRC64;

Query Match 29.6%; Score 628.5; DB 4; Length 353;
Best Local Similarity 43.5%; Pred. No. 2.4e-46;
Matches 134; Conservative 50; Mismatches 105; Indels 19; Gaps 8;

Qy 96 CEPDLRVTSVVTGNNLPDRFKDFLLYLRNYSLLIDQPKCAKKPFLLLAITSKSLTPHF 155
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 51 CPPNHTVSSASL----SLPSRHLFLTYRHCRNFSILL-EPSGCSKDTFLLAITSKSGPHV 106
Qy 156 ARQAIRESGQESN-AGNOTVVRVFL---GOTPPEDNHPDLSMKLFESEKHQDILMW 211
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 107 ERRAAIRSTGVRGSGWARGQLKLVFLLVGAGSAPP-----AQLLAYESREFDDILOW 159
Qy 212 NYRDFENLSLKEVLFRLRWYSTSCPDTFVFVKGDGDDVFNTHILNYSLSKTKAKOLF 271
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 160 DFTEDFNLTKLHLQRWVVAACPAHFMLKGDGDDVFNTHILNYSLSKTKAKOLF 217
Qy 272 IGDVHNAGPHRDKKLYIPEVYVYSGLYSGHLLALRLYHIITDQVHLYP 330
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 218 VGDVIRQALPNRTKVKYFIPPSMYRATHYPPYAGGGYVMSRATVRRLOAIMEDAELFP 277
Qy 331 IDVVYTGMCLOKGLVPEKHGKFTFDIEEK-NKNNNICSYVDMLVHRSKPKQEMIDIWSQ 369
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 278 IDOVFGMCLRLGLSPMHAGKFTGIRRLDPLDPCLYRGILLVHRLSPLEMTWML 337
Qy 390 LQSAHLKC 397
:|||||
Db 338 VTDEGLKC 345

RESULT 15
Q9C0J1 PRELIMINARY; PRT; 378 AA.
ID Q9C0J1;
AC Q9C0J1;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Beta-1,3-N-acetylgalucosaminyltransferase bGn-14.
GN BGNT-4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21264825; Pubmed=11042166;
RA Shiraiishi N., Natsume A., Togayachi A., Endo T., Akashima T.,
RA Yamada Y., Imai N., Nakagawa S., Koizumi S., Sekine S., Narimatsu H.,
RA Sasaki K.;
RT "Identification and Characterization of Three Novel beta1,3-N-
RT Acetylgalucosaminyltransferases Structurally Related to the beta1,3-
RL J. Biol. Chem. 276:3498-3507(2001).
DB EMBL; AB049586; BAB21532.1; -.
DR Genbank; HGNC:15683; B3GNT4.
DR InterPro; IPR002659; Glyco_trans_31.
DR Pfam; PF01762; Galactosyl_T; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 378 AA; 42310 MW; E8EB88665A36057E CRC64;

Query Match 29.6%; Score 628.5; DB 4; Length 378;
Best Local Similarity 43.5%; Pred. No. 2.6e-46;
Matches 134; Conservative 50; Mismatches 105; Indels 19; Gaps 8;

Qy 96 CEPDLRVTSVVTGNNLPDRFKDFLLYLRNYSLLIDQPKCAKKPFLLLAITSKSLTPHF 155
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 76 CPPNHTVSSASL----SLPSRHLFLTYRHCRNFSILL-EPSGCSKDTFLLAITSKSGPHV 131
Qy 156 ARQAIRESGQESN-AGNOTVVRVFL---GOTPPEDNHPDLSMKLFESEKHQDILMW 211
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 132 ERRAAIRSTGVRGSGWARGQLKLVFLLVGAGSAPP-----AQLLAYESREFDDILOW 184
Qy 212 NYRDFENLSLKEVLFRLRWYSTSCPDTFVFVKGDGDDVFNTHILNYSLSKTKAKOLF 271
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 185 DFTEDFNLTKLHLQRWVVAACPAHFMLKGDGDDVFNTHILNYSLSKTKAKOLF 242
Qy 272 IGDVHNAGPHRDKKLYIPEVYVYSGLYSGHLLALRLYHIITDQVHLYP 330
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 243 VGDVIRQALPNRTKVKYFIPPSMYRATHYPPYAGGGYVMSRATVRRLOAIMEDAELFP 302
Qy 331 IDVVYTGMCLOKGLVPEKHGKFTFDIEEK-NKNNNICSYVDMLVHRSKPKQEMIDIWSQ 369
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 303 IDOVFGMCLRLGLSPMHAGKFTGIRRLDPLDPCLYRGILLVHRLSPLEMTWML 362
Qy 390 LQSAHLKC 397
:|||||
Db 363 VTDEGLKC 370

Search completed: October 20, 2003, 23:04:03
Job time : 43 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 22:54:34 ; Search time 12.5 seconds
(without alignments)
1493.568 Million cell updates/sec

Title: US-09-804-357B-2

Perfect score: 2129
Sequence: 1 MSVGRRRYKLLGLMHANVF.....RKQEMIDWSQLSPNLKC 397

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1862	87.5	397	1 B3G7_HUMAN	Q9ny97 h beta-1,3-
2	619.5	29.1	372	1 B3G8_HUMAN	Q9y2a9 h beta-1,3-
3	400	18.8	308	1 B3G5_MOUSE	Q9j167 m beta-1,3-
4	354	16.6	300	1 B3G5_GORGO	Q9n293 g beta-1,3-
5	354	16.6	301	1 B3G5_PANPA	Q9n294 p beta-1,3-
6	353	16.6	310	1 B3G5_HUMAN	Q9y263 h beta-1,3-
7	349	16.4	297	1 B3G5_PANTR	Q9n295 p beta-1,3-
8	314.5	14.6	371	1 B3G4_MOUSE	Q9z0f0 mus musculus
9	310.5	14.6	371	1 B3G4_RAT	O88178 rattus nerv
10	287	13.5	378	1 B3G4_HUMAN	O96024 homo sapien
11	104.5	4.9	1026	1 BGAL_STRTR	P23989 streptococc
12	103.5	4.9	610	1 RCX2_YEAST	P38623 saccharomyc
13	100	4.7	432	1 BFAA_THEMEA	O33833 thermotoga
14	99	4.7	440	1 FENR_ANASO	P21890 anabaena sp
15	99	4.7	517	1 VLI_HPV25	Q02051 human papil
16	98	4.6	1018	1 YC14_METJA	Q58611 methanococc
17	96	4.5	2733	1 RRPB_CVMA5	P16342 murine coro
18	95	4.5	626	1 HTPQ_BUCBP	Q89a93 buchnera ap
19	94.5	4.4	800	1 P127_YEAST	P32606 saccharomyc
20	94	4.4	472	1 KSGT_ARATH	Q96287 arabidopsis
21	93	4.4	440	1 FENR_ANAVA	Q44549 anabaena va
22	92.5	4.3	485	1 SYE2_THETN	O887t1 thermocanae
23	92.5	4.3	995	1 SR13_MOUSE	Q923q2 mus muscu
24	92	4.3	440	1 FENR_ANASP	P58558 anabaena sp
25	89.5	4.2	842	1 DP3A_THEMEA	Q92hg4 thermotoga
26	89.5	4.2	2131	1 YCF2_SPIOL	P09973 spinacia ol
27	89	4.2	468	1 KSGT_BRANA	O04160 brassica na
28	89	4.2	597	1 TTPI_YEAST	P38069 saccharomyc
29	88.5	4.2	249	1 RL7A_SCHPO	P17937 schizosacch
30	88.5	4.2	4466	1 DYHC_TRIGR	P23098 triplosacch
31	87.5	4.1	498	1 GALT_STACA	Q9tgr9 staphylococ
32	87.5	4.1	1024	1 V075_MYCGE	P47321 mycoplasma
33	87	4.1	705	1 PPK_BACHD	Q9kd27 bacillus ha

RESULT 1
B3G7_HUMAN
ID B3G7_HUMAN STANDARD; PRT: 397 AA.
AC Q9NY97; Q9NQ99; Q9NQ90; Q9NUT9;
DI 16-OCT-2001 (Rel. 40, Created)
DI 15-SEP-2003 (Rel. 42, Last sequence update)
DI 15-SEP-2003 (Rel. 42, Last annotation update)
DE Beta-1,3-galactosyltransferase 7 [EC 2.4.1.-] (Beta-1,3-GalTase 7)
DE (Beta3Gal-T7) (b3Gal-T7) (UDP-galactose:beta-N-acetylglucosamine beta-1,3-galactosyltransferase 7) (UDP-Gal:beta-GlcNAc beta-1,3-galactosyltransferase 7) (Beta-3-Gx-T7).
GN B3GNT1 OR B3GALT7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Amado M., Carneiro F., Clausen H.;
RT "Cloning and expression of two beta-1,3-galactosyltransferases:
KI beta3gal-T5 and beta3gal-T6.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RP TISSUE=Urinary bladder;
RC Gromova I., Gromov P., Celis J.E.;
RI "A novel member of beta-1,3-galactosyltransferase family is down
regulated during bladder TCC progression.";
RC Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A. (ISOFORM 1).
RP TISSUE=Brain, and Testis;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Ussid T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Muny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahy J., Heltan E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[4]
RP SEQUENCE OF 25-397 FROM N.A.
RC TISSUE=Placenta;

34 87 4.1 838 1 SYV_MYCPN P75304 mycoplasma
35 87 4.1 1427 1 MES4_DROME Q8mt36 dtosophilla
36 87 4.1 4036 1 RRPL_DUGBV Q66431 dugbe virus
37 86.5 4.1 995 1 MNT3_HUMAN Q9y3m8 homo sapien
38 86 4.0 378 1 SRT3_CANAL P87207 candida alb
39 85.5 4.0 655 1 YKDA_MYCCA P45615 mycoplasma
40 85.5 4.0 1770 1 YCB9_YEAST P25384 saccharomyc
41 85 4.0 395 1 VSX2_BRARE O42477 brachydanio
42 84.5 4.0 451 1 APEA_THEMEA Q9wyj9 thermotoga
43 84.5 4.0 772 1 CPTM_RAT Q63704 rattus norv
44 84.5 4.0 929 1 RIR1_NEUCR Q9uw15 neuropept
45 84.5 4.0 956 1 SYI_AQUAE O66651 aquifex aeo

QY	121	LYLCRNYSLLDIPDKCAKAPPELLAIKSLIPHPARQOAIRESWGRTNVGNQTVVRVF	180
DB	121	LYLCRNYSLLDIPDKCAKAPPELLAIKSLIPHPARQOAIRESWGRTNVGNQTVVRVF	180
QY	181	LLGKTPPEDNHPDLSMDLKAFESDKHODILMWNVROTFFNLSLKEVLFURWYSTSCPDAEF	240
DB	181	LLGKTPPEDNHPDLSMDLKAFESDKHODILMWNVROTFFNLSLKEVLFURWYSTSCPDAEF	240
QY	241	VFKGDVDFVNTTHILNLYNSLSKSKAKDLFIGDVIHNAHPHRDKKLKYYIPEVFIYGVY	300
DB	241	VFKGDVDFVNTTHILNLYNSLSKSKAKDLFIGDVIHNAHPHRDKKLKYYIPEVFIYGVY	300
QY	301	PPYAGGGFLYSGPALLRLRLYSATSRVRLYPIDDVVTGMCLOKLGVLPEKHKGFRTDIEE	360
DB	301	PPYAGGGFLYSGHGLAURLVHTIDQVHLVPIDDVVTGMCLOKLGVLPEKHKGFRTDIEE	360
QY	361	KNKNKNCISYDLMVLVHRSKPKQEMIDWSQLQSPNLKC	397
DB	361	KNKNKNCISYVDLMVLVHRSKPKQEMIDWSQLQSAHLKC	397
RESULT 2			
ID	B3G8_HUMAN	STANDARD:	PRT; 372 AA.
AC	C9Y2A9;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Beta-1,3-galactosyltransferase 8 (EC 2.4.1.-) (Beta-1,3-GalTase 8)		
DE	(Beta3Gal-T8) (B3Gal-T8) (UDP-galactose:beta-N-acetylglucosamine beta-		
DE	1,3-galactosyltransferase 8) (UDP-Gal:beta-GlcNAc beta-1,3-		
DE	galactosyltransferase 8) (Beta-3-Gx-T8) (Core 1 extending beta-1,3-N-		
DE	acetylglucosaminyltransferase) (Core1-beta3GlcNACT).		
GN	B3GNT3 OR B3GALT8 OR TWEM3.		
OS	homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
PP	[1]		
PP	SEQUENCE FROM N.A.		
RX	MEDLINE=99173880; PubMed=10072769;		
RT	Yokoyama-Kobayashi M., Yamaguchi T., Sekine S., Kato S.;		
RT	"Selection of cDNAs encoding putative type II membrane proteins on the		
RT	cell surface from a human full-length cDNA bank.;"		
RL	Gene 228:161-167(1999).		
RL	[2]		
PP	SEQUENCE FROM N.A.		
RX	MEDLINE=21332592; PubMed=11439191;		
RT	Yen J., Hiraoka N., Petryniak B., Nakayama J., Ellies L.G., Rabuka D.,		
RA	Hindsgaul O., Marth J.D., Lowe J.B., Fukuda M.;		
RT	"Novel sulfated lymphocyte homing receptors and their control by a		
RT	core1 extension beta1,3-N-acetylglucosaminyltransferase.;"		
RL	Cell 105:957-969(2001).		
RL	[3]		
PP	SEQUENCE FROM N.A.		
RA	Jensen M.A., Bennett E.P.;		
RT	"Cloning of a new member of the beta 1,3 Galactosyltransferase family,		
RT	b1,3Gal-T6.;"		
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.		
RL	[4]		
PP	REVIEW.		
RX	MEDLINE=20047730; PubMed=10580128;		
RA	Anado M., Almeida R., Schwientek T., Clausen H.;		
RT	"Identification and characterization of large galactosyltransferase		
RT	gene families: galactosyltransferases for all functions.;"		
RL	Biochim. Biophys. Acta 1473:35-53(1999).		
CC	-1- PATHWAY: Glycosylation.		
CC	-1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).		
CC	-1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics Institute. There are no restrictions on its		

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.)
 CC -----
 DR ENBL: AB015630; BAA76497.1; -
 DR ENBL: AF293973; AAK00849.1; -
 DR ENBL: AJ100847; CAC45044.1; -
 DR Genev: HGNC:13528; B3GNT3.
 DR GO: GO:0005887; C: integral to plasma membrane; TAS.
 DR Pfam: PF01762; Galactosyl_T.1
 DR Transf: Glycosyltransferase; Glycoprotein; Transmembrane;
 KW Signal-anchor; Golgi stack; Multigene family.
 FT DOMAIN 1 10 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 11 31 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 32 372 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 362 362 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 367 367 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 372 AA: 42515 MW: A6E3FE88AFCFC2 CRC64;
 Query Match 29.1%; Score 619.5; DB 1; Length 372;
 Best Local Similarity 38.1%; Pred. No. 6.3e-44;
 Matches 135; Conservative 53; Mismatches 133; Indels 33; Gaps 7;
 QY 50 KPSTPRAYNREQELNRYNPILNRVANQTGELATSPNTSHLSYCEPDSTVMTAVDF 109
 DB 39 QPPAIEAL-----AWPTP-----PTRAPAPCHANTSMVTH-PDF 73
 QY 110 NLPORFDFLLYLCRNYSLID-OPKCAKPKPFLLLAIAKSLIPHFARRQAIRSWGRE 168
 DB 74 ATQPQHVQNFLLYRCHREFLLQDPVPSKCAQPVFLLLVKSSPNRYRRELLRRTWGRE 133
 QY 169 TNYGNTVYRVFLLGKTPPEDNHPDLSMLKFESDKHODILMKNYRDTFFNLSKLVFL 228
 DB 134 RKYRGLQLRLFLVGTASNPHEARKYNRLLEAQTGHIDFHDSEFNLTAKQVFL 193
 QY 229 RWVSTSCPDPAEFVKGGDDVFNTHILNLSLSKSKAKDLFIGDVHINAGPHRDXKLK 268
 DB 194 QMOETRCANASFVLDGDDGVFAHTDMVFLQ--DHDPGRHLFVGLIQNVGPIRAFWSK 251
 QY 289 YVPEVYTCV-YPPYAGGGFLYSGPALLRLYSATSRVHLYPIDDDVTGMCLOKLVGP 347
 DB 252 YVPEVYTONERYPPYCGGGFLLSRFTAAALRRAAHVLDIPIDDVFLGMCLEGLKP 311
 QY 348 EKHGFRITFDIEKNKK----NTCSYIDLMVHSRKPQEMIDWISQSPNLKC 397
 DB 312 ASHGIRTSGVRAPSQHLSFSDPCFYRDLLVHRELPYEMLLMDALNQNLTG 365
 RESULT 3
 B3G5_MOUSE
 ID B3G5_MOUSE STANDARD; PRT: 308 AA.
 AC Q9J167;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Beta-1,3-galactosyltransferase 5 (EC 2.4.1.-) (Beta-1,3-galactase 5)
 DE (BetaGal-T5) (b3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
 DE 1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
 DE galactosyltransferase 5) (Beta-3-Gx-T5) (Stage-specific embryonic
 DE antigen-3 synthase) (SSEA-3 synthase).
 OS Mus musculus (Mouse).
 GN B3GALT5 OR B3GT5.
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=129/SvJ;
 RX MEDLINE=20390006; PubMed=10837462;
 RA Zhou D., Henion T.R., Jungalwala F.B., Berger E.G., Hennet T.;
 FT "The beta1,3-galactosyltransferase beta3GalT-V is a stage-specific
 FT embryonic antigen-3 (SSEA-3) synthase";
 RL J. Biol. Chem. 275:22631-22634(2000).
 CC -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
 CC WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
 CC GLCNAC(BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOLIPID LC3CER AS AN
 CC EFFICIENT ACCEPTOR. ALSO CATALYZES THE TRANSFER OF GAL TO THE
 CC TERMINAL GALNAC UNIT OF THE GLOBOSIDE GB4, THEREBY SYNTHESIZING
 CC THE GLYCOLIPID GB5, ALSO KNOWN AS THE STAGE-SPECIFIC EMBRYONIC
 CC ANTIGEN-3 (SSEA-3).
 CC -!- PATHWAY: Glycosylation.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
 CC -!- TISSUE SPECIFICITY: Expressed in brain and kidney.
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.)
 CC -----
 DR ENBL: AF254738; AAF86241.1; -
 DR MGD: MGI:2136878; B3galT5.
 DR InterPro: IPR002659; Glyco_trans_31.
 DR Pfam: PF01762; Galactosyl_T.1.
 KW Transf: Glycosyltransferase; Glycoprotein; Transmembrane;
 KW Signal-anchor; Golgi stack; Multigene family.
 FT DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 8 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 229 229 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 308 AA: 35964 MW: 789073A3178825B1 CRC64;
 Query Match 18.8%; Score 400; DB 1; Length 308;
 Best Local Similarity 30.9%; Pred. No. 7.6e-26;
 Matches 90; Conservative 64; Mismatches 119; Indels 18; Gaps 7;
 QY 105 AVTDFNLPORFDFLLYLCRNYSLIDOPK-KCAKK-PFLLLAIAKSLIPHFARRQAIR 162
 DB 24 SMDSFRELPLPFVK-----KSHGKFLQIPDIDCKQKPPFLVLLVTSKQLAARMAIR 75
 QY 163 ESWGRETNVGNQTVYRVFLLGKTPPEDNHPDLSMLKFESDKHODILMKNYRDTFFNLSL 222
 DB 76 KTWGRETSVQGOQVTFLLGTSDSTEE---MDATTFSEQHRDIIQKDFKDAYFNLI 131
 QY 223 KEVLFRLRWVSTSCPDPAEFVKGGDDVFNTHILNLSLSKSKAKDLFIGDVHINAGPH 282
 DB 132 KTMGMEMWVYHFCPTATYVMKTDSDMEFNVGVYLTLL--LKKNKTRFTFTGYIKPHDFPI 189
 QY 283 RDKLKYI1PEVFTY-GVYPPYAGGGFLYSGPALLRLYSATSRVHLYPIDDDVTGMCLO 341
 DB 190 RQKFNKPFVSKFEYPPWDRIYPPFCSTGYVFSSDVAIQVNVYVSEVPFIKLEDFVGLCLA 249
 QY 342 KLGLVPEKKHKGFRFTDIEKNKNKNCYSYIDLMVHSRKPQEMIDWISQSLQ 392
 DB 250 KLKTRPEELHTKQTF-FPGGLRFSVCVCFQKIVACHFMKPDOLLYWQALEN 299
 RESULT 4
 B3G5_GORGO
 ID B3G5_GORGO STANDARD; PRT: 300 AA.
 AC Q9N293;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

```
DE Beta-1,3-galactosyltransferase 5 (EC 2.4.1.1-) (Beta-1,3-GalTase 5)
DE (Beta3Gal-T5) (b3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
DE 1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
DE galactosyltransferase 5) (Beta-3-Gx-T5) (Fragment).
GN B3GALT5.
OS Gorilla gorilla gorilla (Lowland gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
OX NCBI_TaxID=9595;
RN [1]
RP SEQUENCE FROM N.A.
RL Liu Y., Saitou N.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
CC WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
CC GLCNAC(BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOLIPID LC3CER AS AN
CC EFFICIENT ACCEPTOR (BY SIMILARITY).
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB041416; BAA94501.1;
DR InterPro: IPR002659; Glyco_trans_31.
DR Pfam: PF01762; Galactosyl_T. 1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Signal-anchor; Golgi stack; Multigene family.
FT DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 8 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 29 >300 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 130 130 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 231 231 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 300 300 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 300 AA; 35104 MW; 1AA11692ED9F06FA CRC64;
Query Match 16.6%; Score 354; DB 1; Length 300;
Best Local Similarity 30.3%; Pred. No. 4.7e-22;
Matches 76; Conservative 56; Mismatches 111; Indels 8; Gaps 4;
Qy 142 PFLLLAIKSLIPHARRQAIRESWGRETNGQTVVVRVFLGKTPPEDNHPDLSMLKFE 201
Db 57 PFLVLLVTSKQLAERMAIROTGWKERTYKQKLTFFLLGTTSSAAETREVDQ---E 112
Qy 202 SDKHQDILMNRYDTEFNLSLKEVLFLRWVSTSCPDAEVFYKGGDDVFVNTHHILNLS 261
Db 113 SQRHGDIIOKDFLDVYNTLTMTMGIEWVHRFCQAAAFVMTKDSDFINVDYLTLL-- 170
Qy 262 LSKXAKDPIGDIIVHAGPHRDKKLYIPEVFT-GVYPYAGGGGLYSGPALLRLY 320
Db 171 LKNRTTRFTGFLKNEFPPIQPFSEKWFVSKSEYFWDRIYPPFCSGTGYVSGDVASQV 230
Qy 321 SATSRVHLYPIDDDVTGMLQKLGVLPEKHGKGFRTFDEEKNKKNCISYIDLMLVHSRKP 380
Db 231 NVSKSPYIKLEDVFGVGLCLERLNIRLELHLSOPTF-PFGGLRFSVCLFRIVACHFIKP 289
Qy 381 QEMIDIWSQLQ 391
Db 290 RTLLDYWALE 300
RESULT 5
B3G5_PANPA STANDARD; PRT; 301 AA.
ID B3G5_PANPA
AC Q9N294;
```

```
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Beta-1,3-galactosyltransferase 5 (EC 2.4.1.1-) (Beta-1,3-GalTase 5)
DE (Beta3Gal-T5) (b3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
DE 1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
DE galactosyltransferase 5) (Beta-3-Gx-T5) (Fragment).
GN B3GALT5.
OS Pan paniscus (Pygmy chimpanzee) (Bonobo).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9597;
RN [1]
RP SEQUENCE FROM N.A.
RL Liu Y., Saitou N.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
CC WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
CC GLCNAC(BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOLIPID LC3CER AS AN
CC EFFICIENT ACCEPTOR (BY SIMILARITY).
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB041415; BAA94500.1;
DR InterPro: IPR002659; Glyco_trans_31.
DR Pfam: PF01762; Galactosyl_T. 1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Signal-anchor; Golgi stack; Multigene family.
FT DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 8 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 29 >301 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 130 130 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 231 231 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 301 AA; 35222 MW; E04CC70F41CB29FD CRC64;
Query Match 16.6%; Score 354; DB 1; Length 301;
Best Local Similarity 30.2%; Pred. No. 4.8e-22;
Matches 76; Conservative 57; Mismatches 111; Indels 8; Gaps 4;
Qy 142 PFLLLAIKSLIPHARRQAIRESWGRETNGQTVVVRVFLGKTPPEDNHPDLSMLKFE 201
Db 57 PFLVLLVTSKQLAERMAIROTGWKERTYKQKLTFFLLGTTSSAAETREVDQ---E 112
Qy 202 SDKHQDILMNRYDTEFNLSLKEVLFLRWVSTSCPDAEVFYKGGDDVFVNTHHILNLS 261
Db 113 SQRHGDIIOKDFLDVYNTLTMTMGIEWVHRFCQAAAFVMTKDSDFINVDYLTLL-- 170
Qy 262 LSKXAKDPIGDIIVHAGPHRDKKLYIPEVFT-GVYPYAGGGGLYSGPALLRLY 320
Db 171 LKNRTTRFTGFLKNEFPPIQPFSEKWFVSKSEYFWDRIYPPFCSGTGYVSGDVASQV 230
Qy 321 SATSRVHLYPIDDDVTGMLQKLGVLPEKHGKGFRTFDEEKNKKNCISYIDLMLVHSRKP 380
Db 231 NVSKSPYIKLEDVFGVGLCLERLNIRLELHLSOPTF-PFGGLRFSVCLFRIVACHFIKP 289
Qy 381 QEMIDIWSQLQ 392
Db 290 RTLLDYWALEN 301
RESULT 6
```

B3G5 HUMAN
 ID B3G5 HUMAN STANDARD: PRT: 310 AA.
 AC Q9V2C3: Q9V2C3; Q9PIX6; Q9PIX7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Beta-1,3-galactosyltransferase 5 (EC 2.4.1.-) (Beta-1,3-GalTase 5)
 DE (Beta3Gal-T5) (b3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-galactosyltransferase 5) (Beta-3-Gx-T5).
 GN B3GALT5
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colorectal adenocarcinoma;
 RX MEDLINE=99230269; PubMed=10212226;
 RA Isshiki S., Itoyachi A., Kudo T., Nishihara S., Watanabe M.,
 RA Kubota T., Kitajima M., Shiraiishi N., Sasaki K., Andoh T.,
 RA Narimatsu H.;
 RT "Cloning, expression, and characterization of a novel UDP-galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase (beta3Gal-T5) responsible for synthesis of type 1 chain in colorectal and pancreatic epithelia and tumor cells derived therefrom.";
 RL J. Biol. Chem. 274:12499-12507(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99337698; PubMed=10406968;
 RA Zhou D., Berger E.G., Henne T.;
 RT "Molecular cloning of a human UDP-galactose:GlcNAc beta1,3GalNAc beta1,3 galactosyltransferase gene encoding an O-linked core3-elongation enzyme.";
 RL Eur. J. Biochem. 263:571-576(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Amado M., Carneiro F., Clausen H.;
 RT "Cloning and expression of two beta-1,3-galactosyltransferases: beta3Gal-T5 and beta3Gal-T6.";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
 RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
 RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudo J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordliek G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecher H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Lehrach H., Reinhardt R., Vaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN [5]
 RP SEQUENCE OF 1-298 FROM N.A.
 RA Liu Y., Saitou N.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP REVIEW.
 RX MEDLINE=20047730; PubMed=10580128;
 RA Anado M., Almeida R., Schwientek T., Clausen H.;
 RT "Identification and characterization of large galactosyltransferase gene families: galactosyltransferases for all functions.";
 RL Biochim. Biophys. Acta 1473:35-53(1999).
 CC -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN GLCNAC(BETA1,3)GALNAc STRUCTURE. CAN USE GLYCOLIPID LC3CER AS AN EFFICIENT ACCEPTOR.

CC -!- PATHWAY: Glycosylation.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN STOMACH, JEJUNUM, COLON, PANCREAS, SMALL INTESTINE, TESTIS AND GASTROINTESTINAL AND PANCREATIC CANCER CELL LINES. HARDLY DETECTED IN LUNG, LIVER, ADRENAL GLAND AND PERIPHERAL BLOOD LEUKOCYTES.
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AB020337; BAA77664.1; -;
 CC EMBL: AF145784; BAF07880.1; -;
 CC EMBL: AJ006078; CAB91547.1; -;
 CC EMBL: AL163280; CAB90446.1; -;
 CC EMBL: AB041412; BAA94497.1; -;
 CC EMBL: AB041413; BAA94498.1; -;
 CC Genes: HGNC:920; B3GALT5.
 CC MIM: 604066; -;
 CC GO: GO:0008499; F-UDP-galactose beta-N-acetylglucosamine beta. . .; TAS.
 CC GO: GO:0008486; P:protein amino acid glycosylation; TAS.
 CC InterPro: IPR002659; Glyco_trans_31.
 CC Pfam: PF01762; Galactosyl_T; 1.
 CC Signal-Anchor: Glycosyltransferase; Glycoprotein; Transmembrane;
 CC Signal-anchor: Golgi stack; Multigene family.
 CC DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 8 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 CC DOMAIN 29 310 LUMENAL, CATALYTIC (POTENTIAL).
 CC CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CONFLICT 23 23 F -> Y (IN REF. 3).
 CC CONFLICT 26 26 Y -> N (IN REF. 3).
 CC CONFLICT 42 42 G -> V (IN REF. 5; BAA94497).
 CC CONFLICT 85 85 M -> T (IN REF. 3 AND 5; BAA94497).
 CC CONFLICT 114 114 Q -> R (IN REF. 5; BAA94497).
 CC CONFLICT 234 234 K -> E (IN REF. 5; BAA94497).
 CC CONFLICT 277 277 L -> R (IN REF. 5; BAA94497).
 CC CONFLICT 293 297 LQYQ -> WTGR (IN REF. 5; BAA94498).
 CC SEQUENCE 310 AA; 36189 MW; 4DD7A19E3E648AA9 CRC64;
 Query Match 16.6%; Score 353; DB 1; Length 310;
 Best Local Similarity 30.2%; Pred. No. 6e-22;
 Matches 76; Conservative 57; Mismatches 111; Indels 8; Gaps 4;
 QY 142 PFLLLAIAKSLIPHFAPROAIRSWGRETNGVQTVVRVFLGKTPEDNHPLSDMLKFE 201
 DB 57 PFLVLLVTSKQLAEIRMAIRQTKRMVKQKQLKFLGLTGTSSAAETKEVDQ----E 112
 QY 202 SDKKHDIIMNVRDTEFNLSLKEVLFLRWVTSQDAEFVKGDVFNTHHILNLS 261
 DB 113 SQRHGDILOKFLDYVYNTLTATMGIEVHRCQAFVMTKDSMDINVDYCTELL-- 170
 QY 262 LSKSKAKDLFIGDVHINAGPHRDKLKYIPEVFT-GVYPPYAGGGFLYSGPALLRL 320
 DB 171 LKKNRTTRFTFTGLKLNFEPIRQPFSEKFKVSKSEYPMWRYPYPPFCGTVSGDVASQV 230
 QY 321 SATSRVHLYPIDDVYTGMCLOKLGVLPEKHKGFRTDIEKNKKNCYSIDLMLVHSRKP 380
 DB 231 NVSKSVYTKLEDVFGVGLCLERLNIRLEELHSQPTF-PFGGLRFSVCLFRRIVACHFIKP 289
 QY 381 QEMIDIWSQLQS 392
 DB 290 RTLDYQWALEN 301
 RESULT 7


```

DR EMBL: M63636; AAA63533.1; .
DR PIR: A49750; A49750.
DR HSP: P00722; IBGL.
DR InterPro: IPR006101; Glyco_hydro.2.
DR InterPro: IPR006102; Glyco_hydro.2Iq.
DR InterPro: IPR006104; Glyco_hydro.2Sg.
DR InterPro: IPR006103; Glyco_hydro.2TIM.
DR InterPro: IPR004200; Glyco_hydro.42C.
DR InterPro: IPR004199; Glyco_hydro.42N.
DR Pfam: PF02930; Bgal_small_C.1.
DR Pfam: PF02929; Bgal_small_N.1.
DR Pfam: PF00703; Glyco_hydro.2.1.
DR Pfam: PF02836; Glyco_hydro.2.C.1.
DR Pfam: PF02837; Glyco_hydro.2.N.1.
DR PRINTS: PR00132; GLYDRLASE2.
DR PROSITE: PS00719; GLYCOSYL_HYDROL_F2.1.1.
DR PROSITE: PS00608; GLYCOSYL_HYDROL_F2.2.1.
DR Hydrolase: Glycosidase.
FT ACT_SITE 458 458 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 546 546 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 1026 AA; 117045 MW; B453AFAD68F08C98 CRC64;

Query Match 4.9%; Score 104.5; DB 1; Length 1026;
Best Local Similarity 22.9%; Pred. No. 1.1;
Matches 61; Conservative 42; Mismatches 82; Indels 81; Gaps 15;

QY 147 AIKSLIFHFAARQAIRESWGRETWGNQTVVRVELLCK-----TPEDNHPDLSDM 197
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 120 AVSYVYHFTLNDALKDKKVFISFGVATSFVWNGFNFGVSEDSFTSEF---EISDY 176
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 198 LKFESDKHQDILMNYRDTFENLSIKVFLRWYSTSCPDAAEFVKGDVDVEVNTTHILN 257
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 177 L-VEGDNKLAVVYSTA-----SWL-----EDQFRLVGIIFR 210
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 258 --YINLSKSKAKDLFI-GDV-----IHNAGPHRDKLKYIPEVFTGYVY-- 300
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 211 DVLYAIKPVHVDLFVKGVDYDTQKAGQIDLDKTVGDEYEDKKIKYVLSD--YEGIVTE 268
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 301 -PPVAGGGGLYSGPALRL--YSATS-----RVHLYP-DDVYTCMCLQKLGIVPEKHGKF 353
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 269 GDASVNGDGLSVLENLKIKPWASPKLYDLILHLVDDDD-----QVVEVFPVK-VGF 321
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 354 RFDIEKNKKNICSYIDLMLVHSRK 379
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 322 RREIKDK-----LMLLNGKR 337
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 12
RCK2_YEAST
ID RCK2_YEAST STANDARD; PRT; 610 AA.
AC P38623; Q02532; Q06557;
DT 01-OCT-1994 (Rel. 30, Created)
DI 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Serine/threonine-protein kinase RCK2 (EC 2.7.1.37) (CAM kinase-like
protein kinase CLK1).
GN RCK2 OR CLK1 OR CMK3 OR YLR248W OR L9672.6.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB320;
RX MEDLINE=94156178; PubMed=8112585;
RA Dahlkvist A., Summerhagen P.;
RT Two novel deduced serine/threonine protein kinases from
RL Saccharomyces cerevisiae.
RL Gene 139:27-33(1994).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=YNN 214;
RX MEDLINE=97094707; PubMed=8939941;

```

```

RA Melcher M.L., Thorner J.;
RT "Identification and characterization of the CLK1 gene product, a
RT novel Cam kinase-like protein kinase from the yeast Saccharomyces
RL cerevisiae.";
RL J. Biol. Chem. 271:29958-29968(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=97313267; PubMed=9169871;
RA Johnston M., Hillier L., Riles L., Albertmann K., Andre B., Ansorge W.,
RA Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,
RA Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,
RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz P.,
RA Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhasselt P.,
RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hohnsels J.D.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RL Nature 387:87-90(1997).
CC -!- CATALYTIC ACTIVITY: ATP + a protein -> ADP + a phosphoprotein.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- PTM: AUTOPHOSPHORYLATED.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC CAMK SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X71065; CAAS0389.1; .
DR EMBL: U23464; AAA64421.1; .
DR EMBL: U20865; BAB67392.1; .
DR PIR: S59394; S59394.
DR HSP: Q63450; IAO6.
DR SGD: S0004238; RCK2.
DR GO: GO:0040020; P:regulation of meiosis; IGI.
DR InterPro: IPR000719; Prot_kinase.
DR InterPro: IPR002290; Ser_thr_kinase.
DR Pfam: PF00069; pkinase; 1.
DR ProDom: PD000001; Prot_kinase; 1.
DR SMART: SM00220; S_TKC; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase: Serine/threonine-protein kinase; ATP-binding;
FT Phosphorylation; Calmodulin-binding.
FT DOMAIN 163 478 PROTEIN KINASE.
FT NP_3_ND 201 201 ATP (BY SIMILARITY).
FT BINDING 169 177 ATP (BY SIMILARITY).
FT ACT_SITE 313 313 BY SIMILARITY.
FT DOMAIN 493 506 CALMODULIN-BINDING (POTENTIAL).
FT CONFLICT 109 109 N -> S (IN REF. 3).
FT CONFLICT 188 188 H -> N (IN REF. 3).
FT CONFLICT 233 233 A -> V (IN REF. 3).
FT CONFLICT 328 328 P -> R (IN REF. 3).
FT CONFLICT 456 456 S -> A (IN REF. 3).
FT CONFLICT 569 610 DEQLQNPFQTLDTSTLQRKKVQNDVGTPIPSATIR
E -> MNSWNKICSN (IN REF. 1).
SQ SEQUENCE 610 AA; 68040 MW; DFE7B95E4DF23A8A CRC64;

Query Match 4.9%; Score 103.5; DB 1; Length 610;
Best Local Similarity 20.7%; Pred. No. 0.66;
Matches 84; Conservative 56; Mismatches 151; Indels 115; Gaps 18;

QY 72 PILNRVANOTGLATSPNTSHLSYCEPD---STVMATVTFNPLDRF--KDFLLYLCR 126
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

Db 102 PAIDVHENSEQSLSDPLSDSELSQSEILSIQDDSDDDNMEDEIPEKSFLEQKELI 161
Qy 127 NYSLLIDQPKKCAKPFLLAIKSLIPHFAARQAIRESMGRETNGVQNTVRFVLGKTP 186
Db 162 GYKLINIGEGAFKSVFRAIPAKNSSHEFLTK-----NYRAVAIKVIKKA- 206
Qy 187 PEDNHPOLSDML--KFESDKHQDILMNYRDTFNLKLEVLFLRWVSTSCDP--AEFVF 242
Db 207 -----DLSSINGDHRKXKDKSTKTSSRDQ-----ALKEVALHKTVSAGCSQIVAFIDF 256
Qy 243 KGDD-----DVFVNTHHILNLSLSKSKAKDLFTG-DVIHAG--PHRDKKL 287
Db 257 QETDSYIIIOELTGGEIAGEIVRLTYFSEDLRSRVIKOLALAVKMHSLGVVHRDIKP 316
Qy 288 KYIYPE-VFYTF-----GVYPPYAGGGG----- 309
Db 317 ENLLEFIEFTPSIKPKLRKSDDPQTKADEGIFTGPGVGGGIGIVKLADFLGSKQIFSKN 376
Qy 310 -----LYSGPALLRLYSATSRVHLYPIDDV-YTGMG-----LQKGLGVPEK-HKGF 353
Db 377 TKTCGIVGYTAPVKKVDEHYHKKVDMWGLCCVLYTMWLGFPFFYDEKIDTITEKISRGE 436
Qy 354 RTF-----DIEAKNKNKISYIDLMLVHSRKKPOEMIDINWSQLOSP 393
Db 437 YTLKPMWDEISAGAKNAVSKLEL-----EPSKRYDIDQFLDDP 476

RESULT 13
BFRA_THEME
ID BFRA_THEME STANDARD; PRT; 432 AA.
AC O33833;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Beta-fructosidase (EC 3.2.1.26) (Sucrase) (Invertase).
GN BFRA OR TM1414.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
Raft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.W.,
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
Hedberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
Salberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima.";
RL Nature 398:323-329(1999).
CC -!- FUNCTION: Hydrolysis of sucrose, raffinose, inulin and levan.
CC Specific for the fructose moiety and the beta-anomeric
CC configuration of the glycosidic linkages of its substrates. The
CC enzyme released fructose from sucrose and raffinose, and the
CC fructose polymer inulin is hydrolyzed quantitatively in an exo-
CC type fashion.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -!- MISCELLANEOUS: Has an optimum temperature of 90-95 degrees
CC Celsius. Extremely insensitive to thermo-inactivation.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ001073; CAA04518.1; .
DR EMBL; AE001794; AAD36485.1; .
DR PIR; D72255; D72255.
DR TIGR; TM1414; .
DR InterPro; IPR001362; Glyco_hydro_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR SMART; SM00640; Glyco_32; 1.
DR PROSITE; PS00609; GLYCOSYL-HYDROL_F32; 1.
KW Hydrolase; Glycosidase; Complete proteome.
SQ SEQUENCE 432 AA; 49841 MW; 39F61B2E1BC462B9 CRC64;
Query Match 4.78; Score 100; DB 1; Length 432;
Best Local Similarity 18.90; Pred. No. 0.83; Indels 70; Gaps 10;
Matches 44; Conservative 38; Mismatches 81;
Qy 181 LLGKTPPEDNR-----PDLSDMLKEFESDKHQDILMNYRDTF 217
Db 124 VISKPEEGTHAFPRDKVKNRSGEMRMVLGSGKDEKIGRVLLYTS-----DLFHKVYEGVI 180
Qy 218 F-NLSLKEVLEFLRWVSTSCDPAEFVKGDDVVF-----NTHHILNLSLSKSKAKDLFT 272
Db 181 FEDETTKEI-----ECPD--LVRIGEKDIILYIISTNSLFSMGLKEKGLNVEKR 230
Qy 273 GDVTHNAGPHRDKKLYKVIPEVF-----TGVPYVAGGGGFLYSGPAL 316
Db 231 GLLDHGT-----DFYAAQTFFGTRVVVIGWLSQWLRTGLYPTKRGWNGVSLPR- 281
Qy 317 LRLYSATSRVHLYPIDDVYTGMLCKLGLVPEKRRKGRFTFDEIEKKNKNCYSY 369
Db 282 -ELYVENNELKVPQDEL---LALRKRKVFETAKSGTFLDVRKNSYEIVCF 330

RESULT 14
FENR_ANASO
ID FENR_ANASO STANDARD; PRT; 440 AA.
AC P21890;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ferredoxin-NADP reductase (EC 1.18.1.2) (FNR).
GN PETH.
OS Anabaena sp. (strain PCC 7119).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=1168;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93344523; PubMed=8343609;
RA Fillat M.F., Flores E., Gomez-Moreno C.;
RT "Homology of the N-terminal domain of the peth gene product from
RT Anabaena sp. PCC 7119 to the CcpD phycobilisome linker polypeptide.";
RL Plant Mol. Biol. 22:725-729(1993).
RN [2]
RP SEQUENCE OF 137-440 FROM N.A.
RX MEDLINE=91088322; PubMed=2124680;
RA Fillat M.F., Bakker H.A.C., Weisbeek P.J.;
RT "Sequence of the ferredoxin-NADP(+) reductase gene from Anabaena PCC
RT 7119.";
RL Nucleic Acids Res. 18:7161-7161(1990).
RN [3]
RP SEQUENCE OF 152-183.
RC STRAIN=1403.46;
RX MEDLINE=88132819; PubMed=312746;
RA Sancho J., Peleato M.L., Gomez-Moreno C., Edmondson D.E.;
RT "Purification and properties of ferredoxin-NADP+ oxidoreductase from
RT the nitrogen-fixing cyanobacteria Anabaena variabilis.";
CC -----

DR	PRINTS; PR00406; CYTB5RDTASE.
DR	PRINTS; PR00371; FPNCR.
DR	ProDom: PD002828; CpcD-like_C; 1.
KW	Oxidoreductase; Flavoprotein; NADP; FAD; Thylakoid; Membrane;
KW	Phycobilisome; 3D-structure.
FT	DOMAIN 1 80
FT	NP_BIND 288 306
FT	CONFLICT 180 180
FT	STRAND 152 153
FT	TURN 154 155
FT	STRAND 156 156
FT	STRAND 158 167
FT	TURN 171 172
FT	STRAND 177 183
FT	TURN 185 186
FT	TURN 193 194
FT	STRAND 196 200
FT	STRAND 204 204
FT	TURN 206 207
FT	STRAND 210 210
FT	STRAND 214 218
FT	TURN 222 227
FT	STRAND 231 237
FT	STRAND 240 241
FT	STRAND 250 251
FT	HELIX 253 259
FT	TURN 260 260
FT	TURN 263 264
FT	STRAND 266 273
FT	TURN 283 284
FT	STRAND 286 292
FT	HELIX 293 295
FT	HELIX 296 307
FT	HELIX 309 314
FT	TURN 316 317
FT	STRAND 323 330
FT	HELIX 333 335
FT	TURN 337 338
FT	HELIX 339 348
FT	TURN 350 352
FT	STRAND 353 359
FT	TURN 360 362
FT	STRAND 364 364
FT	TURN 366 367
FT	STRAND 370 370
FT	HELIX 373 379
FT	TURN 380 380
FT	HELIX 381 388
FT	TURN 389 389
FT	TURN 391 392
FT	STRAND 393 399
FT	TURN 401 402
FT	HELIX 404 416
FT	TURN 417 419
FT	HELIX 422 431
FT	TURN 432 433
FT	STRAND 435 440
SQ	SEQUENCE 440 AA; 48865 MW; 8ELF61D0F09338B6 CRC64;
Query Match	
Best Local Similarity 41.3%; Score 99; DB 1; Length 440;	
Matches 72; Conservative 32; Mismatches 108; Indels 126; Gaps 14;	
Oy	17 ANPFIYLIIVEVSKNSODKNG-----KGGVIIPKEKFKWPPSTPRAYWNREQEKLNRWYN 71
Db	17 SRVEVVEVGMQRNEETDQINPIRKSGSVFI-----RVPYNR-----MN 56
Oy	72 PILNRVANQTGELATSPNTSHLSCEPDSVTMTAVTDNNLPDRFKDFLLYLRCRWYSL 131
Db	57 QEMQRITRLGGKIVTITVTSALQQLNGRTTIAT-VTDASS----- 95
Oy	132 IDQPKKAKPFFLLAIKSLIPHPARQAIRESWGRETNYGNOTVVRVFLLGKTPPE--- 188

```
Db 96 -----ETAKS-----EGNGKAFTKIDSGAKAF--AKPFAEQQL 127
QY 189 ---DNHPDLSMLKFESDKHODILMMNYRDTFFNLSLKEVFLRWVSI SCPDAEFVFKGD 245
Db 128 KKDKNGKNTWQAK---AKHADVPVNLRY-----PNAPFIGK-- 161
QY 246 DDVEVNTFHILNLSUSKSKAKDLFIGDVHINAGPHRDKKLYITPEVYTGTVPPYAG 305
Db 162 --VISN-----EPLVKEGGIGIVQHIFKDLTGSLNKYI--EGQSIGIIPPGVD 205
QY 306 GGGFLYSGPALLRLYSATSRVHLYPIDDDVYTGMCLOKL 343
Db 206 KNG-----KPEKRLRYS:ASTRHGDVDDDKTISLCVROL 239

RESULT 15
VLL_HPV25
ID VLI_HPV25 STANDARD: PRT: 517 AA.
AC Q02051;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Major capsid protein L1.
GN L1
OS Human papillomavirus type 25.
OC Viruses; dsDNA viruses, no RNA stage: Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10609;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94265501; PubMed=8205838;
RA Delius H. Hofmann B.;
RT "Primer-directed sequencing of human papillomavirus types.";
RL Curr. Top. Microbiol. Immunol. 186:13-31(1994).
RN [2]
RP SEQUENCE OF 315-358 FROM N.A.
RX MEDLINE=92407963; PubMed=1326639;
RA Chan S.Y., Bernard H.U., Ong C.K., Chan S.P., Birgit H., Delius H.;
RT "Phylogenetic analysis of 48 papillomavirus types and 28 subtypes and
variants: a showcase for the molecular evolution of DNA viruses.";
J. Virol. 66:5714-5725(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M96283; AAA47022.1; -
DR EMBL; X74471; CAA52529.1; -
DR EMBL; M96289; AAA47028.1; -
DR PIR; S36496; S36496.
DR InterPro: IPR002210; PV_capsid_L1.
DR Pfam: PF00500; late_protein_L1; 1.
DR PRINTS; PR00865; HPVcapsIDL1.
DR ProDom; PD000544; PV_capsid_L1; 1.
KW Coat protein; Late protein.
SQ SEQUENCE 517 AA; 58502 MW; 4FCF5CD5C2B5AE0A CRC64;

Query Match 4.7%; Score 99; DB 1; Length 517;
Best Local Similarity 18.0%; Pred. No. 1.3;
Matches 44; Conservative 35; Mismatches 65; Indels 100; Gaps 8;

QY 35 KNGKGGVLP-----KEKFKKPPSTPRAYNR-----EQEKLNR 68
Db 263 RGGKRGDDIPACQIDEGSKMNAFIPIPPNSSQAQYNLNGSNMYFTVSGSLVSSDAQLFNR 322
QY 69 -----WYNPIILNRVANQTGELATSPNTSHLSYCEPDSVTMTAVTDFNNLP 113
Db 323 PFWLQAOQHNGHNGICWFNLQVTVVDNT-----RNTNFSISINSDGTDVSKITDYS-- 374
```

```
QY 114 DRFKDFLLYLRCRNYSLLIDQPKKCAKPFLLAIAKSLIPHFARROAIRSNGRETNVGN 173
Db 375 QKTEYLHRHVEEYELSLLQLCKVPLKAEIL-----AQINAMN 412
QY 174 QTVRVFLLGKTTPEDNHPDLSMDLKFESDKHODILMMNYRDTFFNLSLKEVFLRWVST 233
Db 413 SNILEEMQLGFGVPAPDN-----SIQDTY-----RYIDSLAT 443
QY 234 SCPD 237
Db 444 RCPD 447

Search completed: October 20, 2003, 23:02:30
Job time : 13.5 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 22:56:05 ; Search time 18.5 Seconds
(without alignments)
2063.728 Million cell updates/sec

Title: US-09-804-357B-2
Perfect score: 2129
Sequence: 1 MSVGRRRVRLGLILMANVF.....RKPOEMDIWSQLSPNLK 397

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.76:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	238	11.2	739	T10648	hypothetical prote
2	218	10.2	507	T18639	hypothetical prote
3	193	9.1	657	T18639	protein T18639.13 [i
4	192.5	9.0	642	D96777	hypothetical prote
5	186.5	8.8	384	T20446	hypothetical prote
6	185.5	8.7	262	T20203	hypothetical prote
7	183	8.6	345	T20031	hypothetical prote
8	181	8.5	357	T20029	hypothetical prote
9	181	8.5	376	T20876	hypothetical prote
10	164.5	7.7	332	T20028	hypothetical prote
11	164.5	7.7	684	T86394	protein T24P13.20
12	157	7.4	325	T24762	hypothetical prote
13	155	7.3	376	T24925	hypothetical prote
14	153	7.2	338	T24743	hypothetical prote
15	151	7.1	395	T86458	probable elicitor
16	150.5	7.1	399	T86453	avr9 homolog F9L11
17	140	6.6	398	T86353	protein F2E2.6 [im
18	135.5	6.4	368	T15096	hypothetical prote
19	130.5	6.1	348	T31918	hypothetical prote
20	130.5	6.1	359	T87937	protein F14B6.6 [i
21	130.5	6.1	385	T20879	hypothetical prote
22	128	6.0	401	T86251	hypothetical prote
23	126.5	5.9	406	T04817	hypothetical prote
24	123	5.8	455	T29555	hypothetical prote
25	122.5	5.8	279	T20443	hypothetical prote
26	122.5	5.8	548	T86456	unknown protein [i
27	120.5	5.7	334	T32256	hypothetical prote
28	118	5.5	349	T24744	hypothetical prote
29	116.5	5.5	404	T86186	hypothetical prote

30	113.5	5.3	409	2	A84733	hypothetical prote
31	113.5	5.3	479	2	D86187	hypothetical prote
32	112.5	5.3	414	2	B96808	protein F28K19.2 [i
33	112.5	5.3	560	2	A86214	hypothetical prote
34	112	5.3	353	2	C96573	protein F12M16.19
35	112	5.3	753	2	T24745	hypothetical prote
36	109.5	5.1	253	2	T22387	hypothetical prote
37	104.5	4.9	1026	2	A49750	beta-galactosidase
38	104.5	4.9	885	2	B70393	hypothetical prote
39	100	4.7	432	2	D72255	beta-fructofuranos
40	99	4.7	374	2	G88799	protein T04B2.3 [i
41	99	4.7	412	2	T24441	hypothetical prote
42	99	4.7	440	1	S33479	ferredoxin-NADP re
43	99	4.7	517	2	S36496	L1 protein - human
44	98	4.6	1018	2	E64451	type I restriction
45	97.5	4.6	610	2	S59394	protein kinase RCK

ALIGNMENTS

RESULT 1

T10648
hypothetical protein T13K14.220 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
C:Accession: T10648
R:Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Meves, H.W.; Mayer, K.F.X.; Le
submitted to the Protein Sequence Database, June 1999
A:Reference number: 216991
A:Accession: T10648
A:Molecule type: DNA
A:Residues: 1-739 <BEV>
A:Cross-references: EMBL:AL080282; GSPDB:GN00062; ATSP:T13K14.220
A:Experimental source: cultivar Columbia; BAC clone T13K14
C:Genetics:
A:Gene: ATSP:T13K14.220
A:Map position: 4
A:Introns: 59/3; 96/1; 358/1; 438/3; 535/3; 578/3; 634/3; 670/3

Query Match		11.2%	Score 238;	DB 2;	Length 739;
Best Local Similarity		22.8%	Pred. No. 1.5e-11;		
Matches		75;	Conservative	65;	Mismatches 133;
				Indels	56;
				Gaps	10;
Qy	81	TGEL-----ATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFKDFLLYLRCRNYSLIDQ	134		
Db	451	TGDDVDSIHATSUSTSHPSFSPKATIEFSEWAPPLPG-----	490		
Qy	135	PKCAKKPF-LLLAIKSLIPFAFRRQAIRESWGRVGNQTVVRFVFLGKTPPEDNHPD	193		
Db	491	-----TPFRLFMGVLSATNHFSEMAVRKTMQCHPSIKSSDVVARFFVALNP-----RKE	540		
Qy	194	LSMDMLKEPSDKHQDILMNNYRDTFFNLKLEFLRWVSTSCPAEYFVKGGDDVFVNTH	253		
Db	541	VNMLKKEAEYFGDIVLPMDRYELVVLTKTATICEFGNVTAP---YIMCDDDTFIRVE	597		
Qy	254	HILYNLSLSKAKDLFIGDVIHNAAGPHRDKLKYIPEVETGVYVPPYAGGGGLYS	313		
Db	598	SILKQIDGVSPSEKS--LYMGNLNLNRHRLPTGKVTWTWEE-WPEAVYPPFANGFYIIS	654		
Qy	314	PALLRLYSATSR--VHLYPIDDVTYTMCLQKLGIVPEKHKGFTFDIEENKNKNCISY--	369		
Db	655	NIAKYIVVQNSRHKLRLFKMEDVSMGLVQEFNNSMQ-----PVEYSNKKFCQYGC	706		
Qy	370	-IDMLVHSRRKPOEMIDWISQLSPNLK	397		
Db	707	TILNYTAHYQSPSQMCLNDNLKGRPOC	735		

RESULT 2

T18639
hypothetical protein B0024.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans


```
C;Date: 15-Oct-1999 *sequence_revision 15-Oct-1999 *text_change 15-Oct-1999
C;Accession: T18639
R;McMurray, A.
Submitted to the EMBL Data Library, April 1996
A;Reference number: Z19001
A;Accession: T18639
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-507
A;Cross-references: EMBL:Z19001; PIDN:CAA94876.1; GSPDB:GN00023; CESP:B0024.3
A;Experimental source: clone B0024
C;Genetics:
A;Gene: CESP:B0024.3
A;Map position: 5
A;Introns: 69/2; 118/2; 143/2; 181/1; 246/2; 281/3; 393/2; 437/3

Query Match      10.2%; Score 218; DB 2; Length 507;
Best Local Similarity 22.8%; Pred. No. 4e-10;
Matches 92; Conservative 70; Mismatches 156; Indels 86; Gaps 15;

QY 58 YWNR--OEKLNWYPIILNR-----VANQTGELAT 86
DB 125 YWDSRYAOKISPTFFPMSHLPMDLPDLYFLDLDQGLIIVECTVVIILENTYSNVT 184
QY 87 SPNTSHLSYCEPDSTVMTATDFNNLPDRKD-----FLLYLRGRNYSYL-LIDOPKK-C 138
DB 185 FLMLDRIRSYTDAPFEVNTNFSFKSHLSKSGSTNOFYHAFKQONHTYQFITVPKQOC 244
QY 139 AKKPELLLAIKSLIPFARRQAIRESWGRETN---VGNQTVVVRVFLGKTPPEDNHPDLS 195
DB 245 SNNTKQITLSTAGNFIRQAIRESWGRETN---VGNQTVVVRVFLGKTPPEDNHPDLS 195
QY 196 DMLKFFSDKHODILMNYRDTIFNLSKEVFLRWVSTSCPAERFVKGDQDVFNTHI 255
DB 301 FAQKEIEFDDMIVTDLYSEELLKVAHLSYKQSCQLADQLKIDDDMAYDMGDS 360
QY 256 LNVLSLSKSKAKDLFIGDVHAGPHRDKLKYIPEVFT--GVYPPVAGGGGFLYSGP 314
DB 361 YRSLEOKKQASINGI-SGIWNKNSPVREKHKRWYPTLYSEKPEPPYIDGPIVLGKN 419
QY 315 ALLRLYSATSRVHLYPIDDYV-TGMCLOKL-----GLVPEKHKGFRTD 357
DB 420 AVPRMLEEAKNYNOMVIEDVFTGVTGKALKIKQINWANHLLRYVIELPSRLK----- 473
QY 358 IEKKNKNCYSIDLML--VHSRK-PQEMIDWISLOSPNLKC 397
DB 474 -----CSKGVPLIYAVHNKGFQMIHDGYQKLK--GWRK 506

RESULT 3
G86397
protein T7N9.18 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 *sequence_revision 02-Mar-2001 *text_change 31-Mar-2001
C;Accession: G86397
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzi,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: G86397
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-657 <STO>
A;Cross-references: GB:AE005172; NID:g8778958; PIDN:AAF79857.1; GSPDB:GN00141
C;Genetics:
A;Gene: T7N9.18
```

A;Map position: 1

```
Query Match      9.1%; Score 193; DB 2; Length 657;
Best Local Similarity 26.7%; Pred. No. 6.7e-08;
Matches 67; Conservative 43; Mismatches 95; Indels 46; Gaps 11;

QY 158 QAIRESWGRETNVGNQTVVVRVFLGKTPPEDNHPDLSKDKHODILMNYRDTF 217
DB 425 RMVVRKSMQKLVRSKVVAREFFVALHARKEVND---LKKEAEYFGDIVIPYMDHY 480
QY 218 FNLSLKEVFLRW-VSTSCPDAAEFVKGDQDVFNTHHILNLYLSLSKSKAND-LFIDGY 275
DB 481 DLVVLKTVAIICEYGVNTVA--AKYVMKCDMDTFRVDVAV---IQAEKVKGRSLYIGNI 535
QY 276 IHNAGPHRDKLKYI---PEVFTGYVPPYAGGGGFLYSGPALLRLYSAT----- 323
DB 536 MFNKHPLRTCKWAVTPEEPEY---YPPYANGPGVILS-----YDVKAFIVDDFEQ 584
QY 324 SRVHLYPDIDVYTGMCLOKLGLVPEKHKGFRTDIEKKNKNCISY---IDLMLVHSRXP 380
DB 585 KRLRFKHEVDVSGMWVVEKFN-----ETRPVAVVHSLAKFCQFCIEDYFTAHYQSP 635
QY 381 QEMIDIWSOLQ 391
DB 536 ROMICMWDKLO 646
```

RESULT 4

```
D96777
hypothetical protein F25A4.23 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 *sequence_revision 02-Mar-2001 *text_change 31-Mar-2001
C;Accession: D96777
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alo,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzi,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tall,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: D96777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-642 <STO>
A;Cross-references: GB:AE005173; NID:g5882743; PIDN:AAD55296.1; GSPDB:GN00141
C;Genetics:
A;Gene: F25A4.23
A;Map position: 1

Query Match      9.0%; Score 192.5; DB 2; Length 642;
Best Local Similarity 24.0%; Pred. No. 7.1e-08;
Matches 62; Conservative 57; Mismatches 114; Indels 25; Gaps 8;

QY 144 LLLAIAKSLIPFARRQAIRESWGRETNVGNQTVVVRVFLGKTPPEDNHPDLSKDKFSD 203
DB 395 IFICILSAGNHFERMAVRKSNQHVLIITSAKVAREFFVAL-----HGRKEVWVLLKKEAE 450
QY 204 KHODILMNYRDTFFNLSLKEVFLRWVSTSCPDAAEFVKGDQDVFNTHHILNLYLSLS 263
DB 451 YFGDIVLVPYMDSYDLVLTVAICEHGAALF-SAKYIMKCDMDTFRVDVAV---IQAEKVKGRSLYIGNI 535
QY 264 KSAKOLFIDGVTHNAGPHRDKK--LKYYIPEVFTGYVPPYAGGGGFLYSGPALLRLYS 321
DB 510 EGRS--LYIGNMNYHKKPLRGKWAFTY---EWPEDYPPYANGPGVILSSDIARFIYD 564
QY 322 ATSR--VHLVPIDDVYTGMCLOKLGLVPEKHKGFRTDIEKKNKNCISY---IDLMLVH 376
DB 565 KFERHKLRLFKMEDVSVGMWVE-----HFKNTPNVDYRHSRFRFCQGCVENYTTAH 616
```

```
QY 377 SRKQEMIDINWQLOSPN 394
Db 617 YQSPRQMICLMDKLRGN 634

RESULT 5
T20446
hypothetical protein E03H4.11 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C:Accession: T20446
R:Dobson, R.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19276
A:Accession: T20446
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-384 <WIL>
A:Cross-references: EMBL:Z81492; PIDN:CAB04032.1; GSPDB:GN00019; CESP:E03H4.11
A:Experimental source: clone E03H4
C:Genetics:
A:Gene: CESP:E03H4.11
A:Map position: 1
A:Introns: 46/1; 76/1; 196/3; 263/2; 305/3; 335/3
C:Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

Query Match 8.8%; Score 186.5; DB 2; Length 384;
Best Local Similarity 20.2%; Pred. No. 1.1e-07;
Matches 79; Conservative 82; Mismatches 133; Indels 97; Gaps 15;

QY 1 MSVGRRRVKLLGIILMANVFIYLIVEYKNSQDKNGKGGVLPKEKFWKPPPTPRAYWN 60
Db 32 LDLGHRSPKLPKIVRVASL-----RRNRSSFHSE-----IPKE-----NCTRSGWL 74

QY 61 REQEKLARWNPILNRVANOTGELATSPNTSHLSYCEPSTVMTAVTDFNNLPDRKDFL 120
Db 75 KD-----STAPETQDFG-----SEFVISFADIQ----- 97

QY 121 LYLRCRNYSLIDQPKKCAKKPFLLLAISKLIPIHFAARRAIRESWGREN---VGNQTV 176
Db 98 -----KNYTLWLPYEIAEASOEKDIIMVASRTDSYARRNIMROTWNKSDSEIVANGRM 152

QY 177 VRFLGKTPEDNHPDLSMDLKFESEKHQDILMNRYEDTFNLSLKEVLFLRWVYSTSCP 236
Db 153 KPLFLVGLTQDYK---MKVMQGEAKLYGDIIVVDMNDYEELTYKSLAILLYGVSKAP 209

QY 237 DAERFVGDDVDVFNTHHILNLSLSKRAKDLFIGDVIRNAGP-----H 282
Db 210 RYQMIKIGKIDEDVI-----FFPQKLTALYEQGII-DATPLCAYGVKIQAGARIF 256

QY 283 RDKKLKYYIPEVFT-GVYPPYAGGGFLYSGPALLRLYSATSRVHLYPIDDYY-TGMCL 340
Db 257 RDKNDRWIVPESSYSCSKFPEYVSGMLVYMTWEAAQQIIKSTKYRDFIQVEDVFLTGILA 316

QY 341 QKLGVLPEKKHGRFTF--DIEEKKNKNCISY 369
Db 317 EDLGISVRNLPKFYKPNIDESKSVDIW 347

RESULT 6
T20203
hypothetical protein C54C8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20203
R:Dobson, R.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19237
A:Accession: T20203
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-262 <WIL>
A:Cross-references: EMBL:Z83102; PIDN:CAB05465.1; GSPDB:GN00019; CESP:C54C8.3

A:Experimental source: clone C54C8
C:Genetics:
A:Gene: CESP:C54C8.3
A:Map position: 1
A:Introns: 19/2; 117/3; 184/2; 226/3

Query Match 8.7%; Score 185.5; DB 2; Length 262;
Best Local Similarity 23.6%; Pred. No. 8.1e-08;
Matches 61; Conservative 58; Mismatches 95; Indels 45; Gaps 10;

QY 127 NYS-LLIDQPKKCAKKPFLLLAISKLIPIHFAARRAIRESWGREN---VGNQTVRVVELL 182
Db 20 NYSWLYLPEIETSOEKDILIIVAGRTDSYARRNIIROTWMKSKANSEIVANGMKPFLFV 79

QY 183 GKTTPEDNHPDLSMDLKFESEKHQDILMNRYEDTFNLSLKEVLFLRWVYSTSCPDAEFV 242
Db 80 GLTPGEYK---MKVMQGEAKLYGDIIVVDMNDYEELTYKSLAILLYGVSKAPRYQMIG 136

QY 243 KGDDOV-----FVNTHHILNLSLSKRAKDLFIGDVIRNAGPDRKDLK 288
Db 137 KIDEDVMFPDKLTLYDQGFIDATPLRIY--GLKMQSGANIF-----RDKTHR 183

QY 289 YVYIPEVFT-GVYPPYAGGGFLYSGPALLRLYSATSRVHLYPIDDYY-TGMCLQKLG- 345
Db 184 WYVPESSYSCSKFPEYVSGMLVYMTWEAAQQIIKSTKYRDFIQVEDVFLTGILAEOLGIS 243

QY 346 ---VPEKHK-----GFRFT 356
Db 244 VKMLPEFYKPSDVGVFFSF 262

RESULT 7
T20031
hypothetical protein C47F8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T20031; T22806
R:McLay, K.
submitted to the EMBL Data Library, December 1997
A:Reference number: Z19212
A:Accession: T20031
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-345 <WIL>
A:Cross-references: EMBL:AL009246; PIDN:CAAL15841.1; GSPDB:GN00019; CESP:C47F8.3
A:Experimental source: clone C47F8
R:Kershaw, J.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19617
A:Accession: T22806
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-345 <WIL>
A:Cross-references: EMBL:Z81553; PIDN:CAB04502.1; GSPDB:GN00019; CESP:C47F8.3
A:Experimental source: clone F56H6
C:Genetics:
A:Gene: CESP:C47F8.3
A:Map position: 1
A:Introns: 23/1; 49/1; 169/3; 236/2; 278/3; 308/3
C:Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

Query Match 8.6%; Score 183; DB 2; Length 345;
Best Local Similarity 22.5%; Pred. No. 1.9e-07;
Matches 83; Conservative 71; Mismatches 137; Indels 78; Gaps 16;

QY 48 FWKPPSTPRAYWNREQEKLNRWNPILNRVANOTGELATSPNTSHLSYCEPSTVMTAVT 107
Db 24 FWKLPQNPKSSFVLEERCVQSGWN--ISTLSRPNIDFGSS-----FIV 64

QY 108 DFNNLPDRKDFLLYLRCRNYSLIDQPKKCAKKPFLLLAISKLIPIHFAARRAIRESWG 167
Db 65 SFANIHSQK--WYIL-----PKMKNVTKHDIILMLVYVSKTKNFARRNVLRSWTWN 113
```


QY 254 HILNYLSLSKSKARDLFGDVIHNAHPHRDKK-LKYIPIPEVY--IGVYPPYAGGGGFL 310
 Db 177 KLIPULDDCKVIDPDAAAFYGLLKEGEPVIRKKDAHWYVPDYAYNCTG-YPAYYAGPFYL 235
 QY 311 YSGPALLRLYSATSRVHLYPIDD-VYTGMCLOKGLVPEK---HKGFRFTFOIEEKNKNI 366
 Db 236 ATRKAAKLVLFKTFKQFNPWTVEDSLITGILANDLG-IPRKNLEHVRHYDIDQNESKEI 294
 QY 367 CSYIDLMLVHSRK 379
 Db 295 LAW-----HSFK 301

RESULT 11
 F86394
 protein T24P13.20 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: F86394
 R:Theologian, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, I.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Hulzar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 C:Accession: F86394
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-684 <STO>
 A:Cross-references: GB:A6005172; NID:g9295733; PIDN:AAF87039.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: T24P13.20
 A:Map position: 1

Query Match 7.7%; Score 164.5; DB 2; Length 684;
 Best Local Similarity 22.1%; Pred. No. 1.6e-05;
 Matches 64; Conservative 53; Mismatches 102; Indels 71; Gaps 13;

QY 144 LLLATKSLIPHFAARQAIRESGRETNV-GNOTVVRFL--GKTPPEDNHPDLSMLKFE 201
 Db 413 LVIGVFSTANFKRMARVETWQTDVRSRVARVFFVGLKSP-----LVNLE 462

QY 202 SDKHODILMNVRDFFNLSKEVL-----FLRWVS-----TS 234
 Db 463 -----LWNEARTYGDVQLMPFVDYVSLISWKTALICIFGLSLYCFASITSYFKQTE 514

QY 235 CPDAEFVFGDDVFNTHHILNYLSLSKSKAKDLFGDVIHNAHPHRDKKLYIPEV 294
 Db 515 VDSAKFIMKTDDAFVDEVLSSMTNTRG---LIYGLNSDSQPIRNPDSKWYSYE 572

QY 295 FY-----TGVPYPPYAGGGFLYS---GPALLRLYSATSRVHLYPIDDVTGMCLOKL- 343
 Db 573 FLLKKQEMPEEYPPWAGPGYINSRDIAESVGKLFK-EGNLKMKFLEDVAMGWIABLT 631

QY 344 --GLVPEKHGRTFDIEEKNKNTCSYIDLMLVHSRKPQEMIDWSOLO 391
 Db 632 KHGLEPHVENDGRIISDGCKD-----GVV---VAHYQSPAEMTCLWRKYQ 673

RESULT 12
 T24762
 hypothetical protein T09F5.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T24762
 R:Morimore, B.
 submitted to the EMBL Data Library, December 1996
 A:Reference number: Z19933

A:Accession: T24762
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-325 <WIL>
 A:Cross-references: EMBL:Z83239; PIDN:CAB05807.1; GSPDB:GN00023; CESP:T09F5.1
 A:Experimental source: clone T09F5
 C:Genetics:
 A:Gene: CESP:T09F5.1
 A:Map position: 5
 A:Introns: 19/1; 42/1; 155/3; 223/2; 265/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

Query Match 7.4%; Score 157; DB 2; Length 325;
 Best Local Similarity 23.4%; Pred. No. 2.5e-05;
 Matches 64; Conservative 47; Mismatches 102; Indels 60; Gaps 11;

QY 100 STYMTAVTDF-NLUPDRFKDF-----LLYLRCRNYSLIDOPKKCAKPKPELLLAIKSLIP 153
 Db 42 STLLNSRDRGSHYEISFADIQGSFEWLYL-----PKFELNNPELLIATSRPD 90

QY 154 HFARRQAIRESW-GRETNVGNQTVRVVFLLGKTPPEDNHPDLSMDMLKFESDKHQDILMN 212
 Db 91 DFSRRNAIRKTMNOKTN-----QITSEFMYGLSSKTD--KVRDIVMREAELYRDIWVTS 144

QY 213 YRDTFFNLSLKEVFLRWVSTSCPDAEFVEKGGDDV-----FVNTHHIL 256
 Db 145 LEDSYTKLAPKTLISILLYAVSKVPSAQLIGRVGDVLFPPNLFQSLDKDNYFINTNNS 204

QY 257 NYLNSLSKSKAKDLFGDVIHNAHPHRDKK---YIPEVFTGVVPPYAGGGFLYS 313
 Db 205 IY-----GYIAEKGKPTTSCKCKSRNLF--FKCSNYLSLSGPFELLTR 248

QY 314 PALLRLYSATSRVHLYPIDD-VYTGMCLOKGL 345
 Db 249 PAAEKLNLASKHRDFHOIDDQLITGOMADDAV 281

RESULT 13
 T24925
 hypothetical protein T15D6.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
 C:Accession: T24925
 K:Dobson, R.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19956
 A:Accession: T24925
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-376 <WIL>
 A:Cross-references: EMBL:Z83125; PIDN:CAB05616.1; GSPDB:GN00019; CESP:T15D6.5
 A:Experimental source: clone T15D6
 C:Genetics:
 A:Gene: CESP:T15D6.5
 A:Map position: 1
 A:Introns: 59/1; 87/1; 207/3; 274/2; 316/3; 339/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

Query Match 7.3%; Score 155; DB 2; Length 376;
 Best Local Similarity 21.5%; Pred. No. 4.4e-05;
 Matches 58; Conservative 56; Mismatches 100; Indels 56; Gaps 11;

QY 144 LLLAIAKSLIPHFAARQAIRESGRETNVGNQTVR-----VELLKGTPPEDNHPDLSM 197
 Db 128 ILMIVASRPGSVSRKKVLRKTW---MNKANSKIIRNGRMQVLELVGMVAGD---RDLMA 181

QY 198 LKFECDKHODILMNVRDFTFNLKLEVLFLRWVSTSCPDAEFVEKGGDDVVF----- 250
 Db 182 VKAEASFGDIIVNNLEDYDNLNLPFKVLSLLLTGTHKASDFKIIGKIDDDIFFPDRLRTP 241

QY 251 -----NTHHILNYLSLSKSKAKDLFGDVIHNAHPHRDKKLYIPEVFTGV-Y 300
 Db 242 LLDENVIDSSSYIYGL-----SQDELVY-----RNETKPMYVPEYATNCTKY 286

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 23:00:41 : Search time 12.5 Seconds
(without alignments)
1343.793 Million cell updates/sec

Title: US-09-804-357B-2

Perfect score: 2129

Sequence: 1 MSVGRRRVKKLGLIMANVF.....RKPEMIDWSQLQSPNLKC 397

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	635.5	29.8	378	4	US-09-482-180A-2
2	578	27.1	397	4	US-09-459-133-2
3	565	26.5	389	4	US-09-459-133-13
4	480.5	22.6	326	2	US-09-055-097-6
5	347.5	16.3	325	2	US-09-055-097-5
6	308.5	14.5	331	4	US-09-996-243-209
7	287	13.5	378	2	US-09-055-097-1
8	142	6.7	372	1	US-08-207-904-10
9	103	4.8	363	4	US-09-464-035A-1
10	103	4.8	363	4	US-09-849-562A-1
11	103	4.8	363	4	US-09-849-031A-1
12	101	4.7	350	4	US-09-464-035A-11
13	99.5	4.7	363	4	US-09-464-035A-3
14	97	4.6	363	4	US-09-464-035A-5
15	88	4.1	624	2	US-08-756-317-9
16	87.5	4.1	1024	3	US-09-091-117-5
17	84	3.9	379	4	US-09-328-352-5405
18	83	3.9	476	4	US-09-134-001C-4117
19	82	3.9	409	3	US-09-023-809B-3
20	81.5	3.8	912	2	US-08-951-871-2
21	81.5	3.8	2165	1	US-08-514-975B-2
22	81.5	3.8	2165	5	PCT-US95-12507-2
23	80.5	3.8	990	1	US-08-232-540-2
24	80.5	3.8	990	1	US-08-428-949A-2
25	80.5	3.8	990	1	US-08-428-948A-2
26	80.5	3.8	990	2	US-08-428-946-2
27	80.5	3.8	990	5	PCT-US95-04656-2

28	80.5	3.8	1013	1	US-08-233-008A-8	Sequence 8, Appli
29	80	3.8	784	4	US-09-982-308B-23	Sequence 23, Appli
30	80	3.8	1138	1	US-07-973-320-2	Sequence 2, Appli
31	80	3.8	1138	1	US-07-973-320-4	Sequence 4, Appli
32	79.5	3.7	320	4	US-09-134-001C-4096	Sequence 4096, Ap
33	79.5	3.7	389	4	US-09-464-035A-7	Sequence 7, Appli
34	79.5	3.7	410	1	US-08-792-283A-8	Sequence 8, Appli
35	79.5	3.7	410	2	US-09-105-908-8	Sequence 8, Appli
36	79.5	3.7	410	3	US-09-271-713-8	Sequence 8, Appli
37	79.5	3.7	610	3	US-08-484-661A-33	Sequence 33, Appli
38	79.5	3.7	610	3	US-08-484-661A-35	Sequence 35, Appli
39	79.5	3.7	610	3	US-08-656-664-33	Sequence 33, Appli
40	79.5	3.7	610	3	US-08-656-664-35	Sequence 35, Appli
41	79.5	3.7	610	3	US-08-656-664-54	Sequence 54, Appli
42	79.5	3.7	610	5	PCT-US96-09641-33	Sequence 33, Appli
43	79.5	3.7	610	5	PCT-US96-09641-35	Sequence 35, Appli
44	79.5	3.7	610	5	PCT-US96-09641-54	Sequence 54, Appli
45	79.5	3.7	906	1	US-08-220-151-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1

US-09-482-180A-2
; Sequence 2, Application US/09482180A
; Patent No. 6361985
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Yamamoto, Gayle
; APPLICANT: Gao, Zeren
; APPLICANT: Whitmore, Theodore E.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING
; FILE REFERENCE: 98-80
; CURRENT APPLICATION NUMBER: US/09/482.180A
; CURRENT FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/115,721
; PRIOR FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-482-180A-2

Query Match	29.8%;	Score	635.5;	DB 4;	Length	378;			
Best Local Similarity	39.5%;	Pred.	No. 2.1e-60;						
Matches	146;	Conservative	52;	Mismatches	117;	Indels	55;	Gaps	11;
Qy	40	GVIIPEKEWKPSTPRA---YNNREQEKLNRWNPILNRVANOTGELATSPNTSHLSYC	96						
Db	44	GCLEFLRAAKPAGDPTAHQPFW-----APPIPRHSC	76						
Qy	97	EPDSVTMTAVTDENLPDRKDFLLYLCRNYSLLIDOPKKCAKKPFLLLAISKLIPIHFA	156						
Db	77	PNHVTSSASL---SLPSRHLFLTYRHCNFSILL-EPSCGSKDTFLLAISKQPGHVE	132						
Qy	157	RQAIRESWGRETNYG---NQTVRVFLL---GKTPPEDNHPDLSMDLKFESDKHDI	209						
Db	133	RRAAIRSTWGR---VGGWARGQLKVELLVAGSAPP-----AQLLAYESREFDIL	182						
Qy	210	MNRYDRTFNLSLKEVFLRWVTSCTPOAEFVFGDDDFVNTHTHILNLSLSKSAKD	269						
Db	183	QWDFTEDEFNLTKELHLQRWVVAACPOAHFMKLGDDDFVHVNPVLEFLDGM--DPAQD	240						
Qy	270	LFIGDVHINAGPHRDKLKYIPEVFTGV-YPPVAGGGFLYSGPALLRLYSATSRVHL	328						
Db	241	LVUGDVIOALPNRTKVKYFIPPSMYRATHYPPVAGGGYVMSRATVRQLQAIMEDAEL	300						
Qy	329	YPIDDDYTGMCLOKLGVLPEKHKGFRFTDIEEK-NKNICSYIDLMLVHSRKRPOEMIDIW	387						


```

; FILING DATE: Filed Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Certone, Michael C.
; REGISTRATION NUMBER: 39,112
; REFERENCE/DOCKET NUMBER: PF-0490 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 326 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 2745735
; US-09-055-097-6

Query Match 22.6%; Score 480.5; DB 2; Length 326;
Best Local Similarity 33.7%; Pred. No. 1.1e-43;
Matches 112; Conservative 64; Mismatches 127; Indels 29; Gaps 9;

QY 69 WYNPILNRVNOTGELATSPNTSHLSYCEPDSVTMTAVTDNFNLDPREFDLYLRCRNY 128
DB 21 WYLSITRPTSYTG----SKPFSHLTVARKNFT-----FGNIRTR-----PINPHSF 63
QY 129 SLLIDOPKKCAK--PFLLLAIKSLIPHFAARQAIRESWGRETNGVQTVVRFLLGKTPP 187
DB 64 EFLINEPKCEKNIPFLVLIISTHKEFDARQALRETWGDNENPKGIKATLELLGK--- 120
QY 188 EDNHPDLSMLKFESDKHODILMWNRYRDTFENLSLKEVLFRLWVSTSCPDAEVFVKGDOD 247
DB 121 -NADPVLNQMWQEQSFHDIIEDFIDSYHNLTLTLMGMWVATFCSKAKYVMKTDSD 179
QY 248 VFVNTHLNLNLSLSKSKAKDLFIGDVIHNAIPHROKKLKIYIPEVY--TGVIYPYAGG 306
DB 180 IFVNMNLVYKLLKPKSTKPRRYFTGYVI--NGGPIRDVRSKWMYPRDLYPDNSYPPFCSG 238
QY 307 GGFLYSGPALLRLYSATSRVHLYPIDOVYTGMCLOKGLVPEKHKCFRTDIEKNKNI 366
DB 239 TGIYFSADVAILYKTSLHTFLLHLEDVYVGLCLRLKLGHPFQNSGFNHW-----KMAYSL 294
QY 367 CSYIDLMLVHSRKPEMIDIMSOLOS--PNLKC 397
DB 295 CRYRRVITVHQISPEEMHRIWINDMSSKKHLRC 326

RESULT 5
US-09-055-097-5
; Sequence 5, Application US/09055097
; Patent No. 595282
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Kari J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: HUMAN OXIDIZED LDL RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; FILING DATE: Filed Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Certone, Michael C.
; REGISTRATION NUMBER: 39,112
; REFERENCE/DOCKET NUMBER: PF-0490 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 326 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 2745735
; US-09-055-097-6

Query Match 22.6%; Score 480.5; DB 2; Length 326;
Best Local Similarity 33.7%; Pred. No. 1.1e-43;
Matches 112; Conservative 64; Mismatches 127; Indels 29; Gaps 9;

QY 69 WYNPILNRVNOTGELATSPNTSHLSYCEPDSVTMTAVTDNFNLDPREFDLYLRCRNY 128
DB 21 WYLSITRPTSYTG----SKPFSHLTVARKNFT-----FGNIRTR-----PINPHSF 63
QY 129 SLLIDOPKKCAK--PFLLLAIKSLIPHFAARQAIRESWGRETNGVQTVVRFLLGKTPP 187
DB 64 EFLINEPKCEKNIPFLVLIISTHKEFDARQALRETWGDNENPKGIKATLELLGK--- 120
QY 188 EDNHPDLSMLKFESDKHODILMWNRYRDTFENLSLKEVLFRLWVSTSCPDAEVFVKGDOD 247
DB 121 -NADPVLNQMWQEQSFHDIIEDFIDSYHNLTLTLMGMWVATFCSKAKYVMKTDSD 179
QY 248 VFVNTHLNLNLSLSKSKAKDLFIGDVIHNAIPHROKKLKIYIPEVY--TGVIYPYAGG 306
DB 180 IFVNMNLVYKLLKPKSTKPRRYFTGYVI--NGGPIRDVRSKWMYPRDLYPDNSYPPFCSG 238
QY 307 GGFLYSGPALLRLYSATSRVHLYPIDOVYTGMCLOKGLVPEKHKCFRTDIEKNKNI 366
DB 239 TGIYFSADVAILYKTSLHTFLLHLEDVYVGLCLRLKLGHPFQNSGFNHW-----KMAYSL 294
QY 367 CSYIDLMLVHSRKPEMIDIMSOLOS--PNLKC 397
DB 295 CRYRRVITVHQISPEEMHRIWINDMSSKKHLRC 326

RESULT 5
US-09-055-097-5
; Sequence 5, Application US/09055097
; Patent No. 595282
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Kari J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: HUMAN OXIDIZED LDL RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; FILING DATE: Filed Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Certone, Michael C.
; REGISTRATION NUMBER: 39,112
; REFERENCE/DOCKET NUMBER: PF-0490 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1150971
; US-09-055-097-5

Query Match 16.3%; Score 347.5; DB 2; Length 325;
Best Local Similarity 32.0%; Pred. No. 3.2e-29;
Matches 90; Conservative 53; Mismatches 119; Indels 19; Gaps 7;

QY 118 DELLYLRCRNY--LLIDOPKKCAKPPFLLLAIKSLIPHFAARQAIRESWGRETNGVQNT 175
DB 58 DKFAYLRVPTAEVPPVDPAR-----LTMLIKSAGVNSRRREAIRRTWGYEGRFSDVH 111
QY 176 VYRVFLLGKTPEDNHPDLSMLKFESDKHODILMWNRYRDTFENLSLKEVLFRLWVSTSC 235
DB 112 LRRVFLGTA--EDSEKDA----WESREHGDILQADTFDAYFNILATMLGMRWASEQF 165
QY 236 PDAEVFVKGDODVFNTHILNLSLSKSKAKDLFIGDVIHNAIPHROKKLKIYIIP--EV 294
DB 166 NSESEYLFVDDDDYYSAKNVLFGLGRQSHOPELLFAGHVQISPLRHKFSKWVSLLE 225
QY 295 FYTGVIYPYAGGGFLYSGPALLRLYSATSRVHLYPIDOVYTGMCLOKGLVPEKHKGFR 354
DB 226 YPFDRMPYVTAGAFILSQKALRQLYAAASVHLPLFRFDDVYLGIVALKAGISLQHCDDER 285
QY 355 TFDIEKNKKNICSYIDLMLVHS--RKPOEMIDIMSOLOSPN 394
DB 286 ---FHRPAYKGPDSYSSVSIASHEFGDPEENTRVWNECRSAN 323

RESULT 6
US-09-996-243-209
; Sequence 209, Application US/09996243
; Patent No. 6478825
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kijavir, Ivar J.
; APPLICANT: Napier, Mary A.
```

```

; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/055,097
; FILING DATE: Filed Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Certone, Michael C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0490 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1150971
; US-09-055-097-5

Query Match 16.3%; Score 347.5; DB 2; Length 325;
Best Local Similarity 32.0%; Pred. No. 3.2e-29;
Matches 90; Conservative 53; Mismatches 119; Indels 19; Gaps 7;

QY 118 DELLYLRCRNY--LLIDOPKKCAKPPFLLLAIKSLIPHFAARQAIRESWGRETNGVQNT 175
DB 58 DKFAYLRVPTAEVPPVDPAR-----LTMLIKSAGVNSRRREAIRRTWGYEGRFSDVH 111
QY 176 VYRVFLLGKTPEDNHPDLSMLKFESDKHODILMWNRYRDTFENLSLKEVLFRLWVSTSC 235
DB 112 LRRVFLGTA--EDSEKDA----WESREHGDILQADTFDAYFNILATMLGMRWASEQF 165
QY 236 PDAEVFVKGDODVFNTHILNLSLSKSKAKDLFIGDVIHNAIPHROKKLKIYIIP--EV 294
DB 166 NSESEYLFVDDDDYYSAKNVLFGLGRQSHOPELLFAGHVQISPLRHKFSKWVSLLE 225
QY 295 FYTGVIYPYAGGGFLYSGPALLRLYSATSRVHLYPIDOVYTGMCLOKGLVPEKHKGFR 354
DB 226 YPFDRMPYVTAGAFILSQKALRQLYAAASVHLPLFRFDDVYLGIVALKAGISLQHCDDER 285
QY 355 TFDIEKNKKNICSYIDLMLVHS--RKPOEMIDIMSOLOSPN 394
DB 286 ---FHRPAYKGPDSYSSVSIASHEFGDPEENTRVWNECRSAN 323

RESULT 6
US-09-996-243-209
; Sequence 209, Application US/09996243
; Patent No. 6478825
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kijavir, Ivar J.
; APPLICANT: Napier, Mary A.
```


APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: ROY, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zhenin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730PLC13
CURRENT FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090535
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090540
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090542
PRIOR FILING DATE: 1998-06-24


```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/207,904
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/908,242
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lazar, Steven R.
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: CGC 1624
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8615
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 372 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-207-904-10

Query Match 6.7%; Score 142; DB 1; Length 372;
Best Local Similarity 20.5%; Pred. No. 8.5e-07;
Matches 72; Conservative 73; Mismatches 138; Indels 68; Gaps 15;

QY 47 KFWKPPST-----PRAYNRQEKLNRMYPILNKVANQTGELATSPNTSHLSYCEP 98
DB 5 QIWTQPNSEMNNDLVIPAFNHEKLTISRECDPKRLAESNGDIMGVKKTHQATQSL 64
QY 99 DSTVMTAVTDF-----NNLPDRPKDFLLYLRCNYSLLIDQPKCAKPPFLLLAI 148
DB 65 DKSMTLELAIARTROTISHNAKEN-----RASNHT-----TPNKAFIVVG 108
QY 149 KSLIPHARRQAIRESWG-----RETNVGNQIVYRVFLLGKTPPEDNHPDLSMLK 202
DB 109 NTAFSSRRKRSRLRETNWPKGDKLRKLEKEGIVIR-FVIGHSATRGVLDRA--D 155
QY 203 DKHODILMNNYRDTFFNLSLKEVLFRLRWSTSCPDAEFVFKGDDDDVFVNTHHILN 262
DB 166 AQYKDFLRDHVEGVHELSTRLVFS-KAVSIWDADFYVKVDDVHLNLGML---AN 221
QY 263 SKSRKAD-LFTGDVHNAGPRDRKK-LKYIYPEVYTG-----VYPPYAGGGFLYS 316
DB 222 AKYKSKPVYIGCM-KSGPVLSQGVRYEYEPYKFEENKRYFRHATQIYGISRD 279
QY 317 LRLYSATSRVHLPIDDVYTGMCQKLGVLPEKKHGTFTDIEKNNKNC 367
DB 280 SYISINSGLIHYANEDVSLGSWL--IGL-----EVEHVDERSMC 317

RESULT 9
US-09-464-035A-1
; Sequence 1, Application US/09464035A
; Patent No. 6492152
; GENERAL INFORMATION:
; APPLICANT: Canfield et al.
; TITLE OF INVENTION: Core 1 3-Galactosyl Transferases & Methods of Use
; FILE REFERENCE: 5820.534
; CURRENT APPLICATION NUMBER: US/09/464,035A
; CURRENT FILING DATE: 1999-12-15

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WordPerfect 8.0 *Software*
; SEQ ID NO 1
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-464-035A-1

Query Match 4.8%; Score 103; DB 4; Length 363;
Best Local Similarity 20.9%; Pred. No. 0.014;
Matches 67; Conservative 43; Mismatches 102; Indels 108; Gaps 17;

QY 76 RVANOTGELATSPNTSHLSYCEPSTVMTAVTDFNNLPDRFKD-----FLYLRCR 129
DB 34 KVDTPNVLHNDPHARHSD--DNGQNHLEGOMNFNADSSQHKDENTDIAENLYQKVRIL 91
QY 130 LLIDOPKKCAKPPFLLLAIKSLIPHARRQAIRESWGRETNVGNQIVYRVFLLGKT 189
DB 92 WYMTGPQNLEKK-----AKHVATWAQRN-----KVLFMSS---EE 125
QY 190 NHPDLSMDLKPFESDKHQDILMNNYRDTFFNLSLKEVLFRLRWSTSCPDAEFVFK 248
DB 126 NKDFPAVGLKTEGRDQ--LYW-----KTIKAFQYVHEHYLEDADWFLKADDDT 172
QY 249 FVNTHHILNLSLSKSAKDLFIGDVIHNAQPHRDKLKYIPE--VFYTGVPYPPY--- 303
DB 173 YV-----ILDNLRLW-----LSKYDPEEPIYFGRRKFPVKQ 204
QY 304 ---AGGGFLYSGPALLRLYSA--TSR-VHLYPIDDYVTGMCLOKLG----- 345
DB 205 GYMGGAGYVLSKEALKREVDAFKTKCTHSSSIDLALGRMEIMNVAGDSRDTICKE 264
QY 346 -----VPEKH--KGF--RTF 356
DB 265 TFHPFVPEHHLIKGYLPRTF 284

RESULT 10
US-09-849-562A-1
; Sequence 1, Application US/09849562A
; Patent No. 6545123
; GENERAL INFORMATION:
; APPLICANT: Cummings, Richard D.
; APPLICANT: McEever, Rodger P.
; TITLE OF INVENTION: synthetic glycosulfopeptides and methods of synthesis thereo
; FILE REFERENCE: 5820.594
; CURRENT APPLICATION NUMBER: US/09/849,562A
; CURRENT FILING DATE: 2001-09-04
; PRIOR FILING DATE: 09/334,013
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 1
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-849-562A-1

Query Match 4.8%; Score 103; DB 4; Length 363;
Best Local Similarity 20.9%; Pred. No. 0.014;
Matches 67; Conservative 43; Mismatches 102; Indels 108; Gaps 17;

QY 76 RVANOTGELATSPNTSHLSYCEPSTVMTAVTDFNNLPDRFKD-----FLYLRCR 129
DB 34 KVDTPNVLHNDPHARHSD--DNGQNHLEGOMNFNADSSQHKDENTDIAENLYQKVRIL 91
QY 130 LLIDOPKKCAKPPFLLLAIKSLIPHARRQAIRESWGRETNVGNQIVYRVFLLGKT 189
DB 92 WYMTGPQNLEKK-----AKHVATWAQRN-----KVLFMSS---EE 125
QY 190 NHPDLSMDLKPFESDKHQDILMNNYRDTFFNLSLKEVLFRLRWSTSCPDAEFVFK 248
DB 126 NKDFPAVGLKTEGRDQ--LYW-----KTIKAFQYVHEHYLEDADWFLKADDDT 172
```



```
Db 199 KPVYKQYMGSGAGYVLSKEALRRVDFAKTEKTHSSSIEDLALGRMCMEIKVBEAGDSR 258
Qy 346 -----VPEKH--KGF--RTF 356
Db 259 DPTGKETFPFVPEHHLIKGLPKTF 284
```

RESULT 14

```
US-09-464-035A-5
; Sequence 5, Application US/09464035A
; Patent No. 6492152
; GENERAL INFORMATION:
; APPLICANT: Canfield et al.
; TITLE OF INVENTION: Core 1 3-Galactosyl Transferases & Methods of Use
; FILE REFERENCE: 5820.534
; CURRENT APPLICATION NUMBER: US/09/464.035A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: WordPerfect 8.0 *Software*
; SEQ ID NO 5
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-464-035A-5
```

```
Query Match 4.6%; Score 97; DB 4; Length 363;
Best Local Similarity 20.4%; Pred. No. 0.064;
Matches 55; Conservative 41; Mismatches 73; Indels 100; Gaps 15;
Qy 121 LYLRCRNYSLLDQPKKCAKPFLLLAIKSLIPHAFARQQAIRESGRETNVGNQTVRVF 180
Db 83 LYQVKVILCWVMSFQWLEKK-----AKVVKATVQRN-----KVL 119
Qy 181 LLGKTPPDNHPDLSDMLKPSDKHODILMWNRYDFTFNLSLKEVLFRLKVVTS-CPDAE 239
Db 120 FMSS---EENQDFTVGLTKEGREQ--LYW-----KTIKAFQVVDHYLEAD 163
Qy 240 FVFGKDDVDVFNTHHILNYNSLSKSKAKDLFGDVIHNAHPHRDKKLKYYIPE--VFT 297
Db 164 WFMKADDTYV---IVDNLWL-----LSKYNPEQDIYEG 195
Qy 298 GVYPPY-----AGGGGLYSGPALLRLYSA--TSR-VHLYPIDDYVTGMCLOKGLG--- 345
Db 196 RRFEPYKQYMGSGAGYVLSKEALRRVNAFKTEKTHSSSIEDLALGRMCMEIINVEAG 255
Qy 346 -----VPEKH--KGF--RTF 356
Db 256 DSRDTIGKETFPFVPEHHLIKGLPKTF 284
```

RESULT 15

```
US-08-756-317-9
; Sequence 9, Application US/08756317
; Patent No. 5849894
; GENERAL INFORMATION:
; APPLICANT: Clemente, Thomas E.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Mitsky, Timothy A.
; APPLICANT: Stark, David M.
; TITLE OF INVENTION: Improved Rhodospirillum Rubrum
; TITLE OF INVENTION: Poly-B-Hydroxyalkanoate Synthese
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/756,317
; FILING DATE: 25-NOV-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,693
; FILING DATE: 29-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Patterson, Melinda L.
; REGISTRATION NUMBER: 33,062
; REFERENCE/DOCKET NUMBER: MOBT:008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 787-1400
; TELEFAX: (713) 787-1440
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 624 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-756-317-9
```

```
Query Match 4.1%; Score 88; DB 2; Length 624;
Best Local Similarity 20.3%; Pred. No. 1.4; 99; Indels 88; Gaps 12;
Matches 58; Conservative 41; Mismatches 41;
Qy 49 WKPFSTPRAYNNREQEKLNRWYNPILNRVANQTGELATSPNTSHLSYCEPDSTV----- 102
Db 322 WKNPDP--SYGDTGMDDDYVSAYLEVMRDVL---DLTDQKKNVAGYCIAGTTLALIPVV 375
Qy 103 -----MTAVTDFNLLPDRFKDFLLYLRCRNYSLLDQPKKCAKPFLLAI 148
Db 376 LKQGRDORVNAATFTALTDFADQ---GEFTAYLOEDFVSGIEEAARTG-----ILGA 426
Qy 149 KSLIPHAFARQQAIRESGRETNVGNQTVRVFLLGKTPPDNHPDLSDMLKPSDKHODI 208
Db 427 QLMTRTSEFLRANDLVWG-----PAIRSYMGLGETPPA-----FDL 461
Qy 209 LMWNYRDTFFNLSLKEVLFRLWVYSTSCPDAEFVFKGDD-----DYFV-----NT 252
Db 462 LFWNGDGT---NLPGRMAVEYLRGLCQQRNFVKEGDLGMHRLHVGDDVTVPLCAIACET 517
Qy 253 RHLLNLYNS-----LSKSKAKDLFGDVIHNAHAG---PHRDKKLKYY 290
Db 518 DRIAPWKSWSRGIAQMGSRDKTIFLSESGHIAIGIVNPPSKKXYGHY 563
```

Search completed: October 20, 2003, 23:05:18
Job time : 13.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 23:04:10 ; Search time 30 Seconds
(without alignments)
2169.170 Million cell updates/sec

Title: US-09-804-357B-2

Perfect score: 2129

Sequence: 1 MSVGRRRVKLLGILMANVF.....RKQEMIDINSQLOPNLKC 397

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 609560 seqs, 163917102 residues

Total number of hits satisfying chosen parameters: 609560

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PTI_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2129	100.0	397	9	US-09-804-357-2
2	2129	100.0	397	10	US-09-804-006-2
3	2113	99.2	397	14	US-10-109-563-2
4	1979	93.0	365	9	US-09-804-357-4
5	1979	93.0	365	10	US-09-804-006-4
6	1862	87.5	397	9	US-09-804-357-14
7	1862	87.5	397	10	US-09-972-912-2
8	1862	87.5	397	10	US-09-804-006-14
9	1862	87.5	397	14	US-10-109-563-4
10	1725	81.0	365	9	US-09-804-357-15
11	1725	81.0	365	10	US-09-804-006-15
12	656	30.8	352	9	US-09-739-451-4
13	656	30.8	353	9	US-09-739-451-12
14	656	30.8	353	15	US-10-097-065-137
15	654.5	30.7	401	12	US-10-288-252-10

16	647.5	30.4	378	12	US-10-237-496-36	Sequence 36, Appl
17	647.5	30.4	378	12	US-10-242-074-36	Sequence 36, Appl
18	647.5	30.4	378	12	US-10-242-505-36	Sequence 36, Appl
19	647.5	30.4	378	12	US-10-242-574-36	Sequence 36, Appl
20	647.5	30.4	378	12	US-10-243-261-36	Sequence 36, Appl
21	647.5	30.4	378	12	US-10-243-282-36	Sequence 36, Appl
22	647.5	30.4	378	12	US-10-243-402-36	Sequence 36, Appl
23	647.5	30.4	378	12	US-10-243-431-36	Sequence 36, Appl
24	647.5	30.4	378	12	US-10-243-164-36	Sequence 36, Appl
25	647.5	30.4	378	12	US-10-244-972-36	Sequence 36, Appl
26	647.5	30.4	378	12	US-10-197-942-36	Sequence 36, Appl
27	647.5	30.4	378	12	US-10-238-196-36	Sequence 36, Appl
28	647.5	30.4	378	12	US-10-245-013-36	Sequence 36, Appl
29	647.5	30.4	378	15	US-10-245-103-36	Sequence 36, Appl
30	647.5	30.4	378	15	US-10-245-107-36	Sequence 36, Appl
31	647.5	30.4	378	15	US-10-245-143-36	Sequence 36, Appl
32	647.5	30.4	378	15	US-10-245-771-36	Sequence 36, Appl
33	647.5	30.4	378	15	US-10-245-851-36	Sequence 36, Appl
34	647.5	30.4	378	15	US-10-245-883-36	Sequence 36, Appl
35	647.5	30.4	378	15	US-10-237-535-36	Sequence 36, Appl
36	647.5	30.4	378	15	US-10-238-183-36	Sequence 36, Appl
37	647.5	30.4	378	15	US-10-238-283-36	Sequence 36, Appl
38	647.5	30.4	378	15	US-10-238-370-36	Sequence 36, Appl
39	647.5	30.4	378	15	US-10-245-055-36	Sequence 36, Appl
40	647.5	30.4	378	15	US-10-245-147-36	Sequence 36, Appl
41	647.5	30.4	378	15	US-10-245-730-36	Sequence 36, Appl
42	647.5	30.4	378	15	US-10-245-739-36	Sequence 36, Appl
43	647.5	30.4	378	15	US-10-246-210-36	Sequence 36, Appl
44	647.5	30.4	378	15	US-10-239-196-36	Sequence 36, Appl
45	647.5	30.4	378	15	US-10-243-024-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-09-804-357-2
; Sequence 2, Application US/09804357
; Patent No. US20010024808A1
; GENERAL INFORMATION:
; APPLICANT: Whlce, David
; APPLICANT: Zhou, Jiahong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/109001
; CURRENT APPLICATION NUMBER: US/09/804,357
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/195,896
; PRIOR FILING DATE: 1998-11-19
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-804-357-2

Query Match: 100.0%; Score 2129; DB 9; Length 397;
Best Local Similarity 100.0%; Pred. No. 1.6e-208;
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MSVGRRRVKLLGILMANVFYILVEYKNSQDKNGKGVIIIPKEKFKWKPSTPRAYN 60
Db 1 MSVGRRRVKLLGILMANVFYILVEYKNSQDKNGKGVIIIPKEKFKWKPSTPRAYN 60
Qy 61 REQEKLNRWYNILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFKDEL 120
Db 61 REQEKLNRWYNILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFKDEL 120

```
QY 121 LYLCRCNYSLLIDOPKCKAKKFFLLAIAKSLIPHFAKQA:RESWGRETNGNOTVVRVF 180
|||||
Db 121 LYLCRCNYSLLIDOPKCKAKKFFLLAIAKSLIPHFAKQA:RESAGRETNGNOTVVRVF 180
|||||
QY 181 LLGKTPPEONHPDLSMDLKFESDKHODILMWNRYRDTFFNLSLKEVLFLEWVSTSCPDADF 240
|||||
Db 181 LLGKTPPEONHPDLSMDLKFESDKHODILMWNRYRDTFFNLSLKEVLFLEWVSTSCPDADF 240
|||||
QY 241 VFKGDDVFNTHHILNYLNSLSKSKAKDLFGDVIHNAHPHRRDKKLYIPEVFTGVY 300
|||||
Db 241 VFKGDDVFNTHHILNYLNSLSKSKAKDLFGDVIHNAHPHRRDKKLYIPEVFTGVY 300
|||||
QY 301 PPYAGGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKLGVLPEKHGKGFRTFDEE 360
|||||
Db 301 PPYAGGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKLGVLPEKHGKGFRTFDEE 360
|||||
QY 361 KKKKNCISYIDLMLVHSRKPQEMIDWISOLQSPNLKC 397
|||||
Db 361 KKKKNCISYIDLMLVHSRKPQEMIDWISOLQSPNLKC 397
|||||

RESULT 2
US-09-804-006-2
; Sequence 2, Application US/09804006
; Patent No. US20020119517A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/126001
; CURRENT APPLICATION NUMBER: US/09/804,006
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/292,228
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-804-006-2

Query Match 100.0%; Score 2129; DB 10; Length 397;
Best Local Similarity 100.0%; Pred. No. 1.6e-208;
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSVGRRRVKLLGILMMANFYIYLIVEVSKNSQDKNGKGVIIIPKEKFWKPPSTPRAYN 60
|||||
Db 1 MSVGRRRVKLLGILMMANFYIYLIVEVSKNSQDKNGKGVIIIPKEKFWKPPSTPRAYN 60
|||||
QY 61 REQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFKDFL 120
|||||
Db 61 REQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFKDFL 120
|||||
QY 121 LYLCRCNYSLLIDOPKCKAKKFFLLAIAKSLIPHFAKQA:RESWGRETNGNOTVVRVF 180
|||||
Db 121 LYLCRCNYSLLIDOPKCKAKKFFLLAIAKSLIPHFAKQA:RESAGRETNGNOTVVRVF 180
|||||
QY 181 LLGKTPPEONHPDLSMDLKFESDKHODILMWNRYRDTFFNLSLKEVLFLEWVSTSCPDADF 240
|||||
Db 181 LLGKTPPEONHPDLSMDLKFESDKHODILMWNRYRDTFFNLSLKEVLFLEWVSTSCPDADF 240
|||||
QY 241 VFKGDDVFNTHHILNYLNSLSKSKAKDLFGDVIHNAHPHRRDKKLYIPEVFTGVY 300
|||||
Db 241 VFKGDDVFNTHHILNYLNSLSKSKAKDLFGDVIHNAHPHRRDKKLYIPEVFTGVY 300
|||||
QY 301 PPYAGGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKLGVLPEKHGKGFRTFDEE 360
|||||
Db 301 PPYAGGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKLGVLPEKHGKGFRTFDEE 360
|||||
```

```
Db 301 PPYAGGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKLGVLPEKHGKGFRTFDEE 360
|||||
QY 361 KKKKNCISYIDLMLVHSRKPQEMIDWISOLQSPNLKC 397
|||||
Db 361 KKKKNCISYIDLMLVHSRKPQEMIDWISOLQSPNLKC 397
|||||

RESULT 3
US-10-109-563-2
; Sequence 2, Application US/10109563
; Publication No. US20020170075A1
; GENERAL INFORMATION:
; APPLICANT: Leviten, Michael W.
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE COMPRISING
; FILE REFERENCE: BETA-1,3-N-ACETYLLACTOSAMINE DISRUPTIONS
; CURRENT APPLICATION NUMBER: US/10/109,563
; CURRENT FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: US 60/280,706
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-109-563-2

Query Match 99.2%; Score 2113; DB 14; Length 397;
Best Local Similarity 99.2%; Pred. No. 6.9e-207;
Matches 394; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSVGRRRVKLLGILMMANFYIYLIVEVSKNSQDKNGKGVIIIPKEKFWKPPSTPRAYN 60
|||||
Db 1 MSVGRRRVKLLGILMMANFYIYLIVEVSKNSQDKNGKGVIIIPKEKFWKPPSTPRAYN 60
|||||
QY 61 REQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFKDFL 120
|||||
Db 61 REQEKLNRYNPILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFKDFL 120
|||||
QY 121 LYLCRCNYSLLIDOPKCKAKKFFLLAIAKSLIPHFAKQA:RESWGRETNGNOTVVRVF 180
|||||
Db 121 LYLCRCNYSLLIDOPKCKAKKFFLLAIAKSLIPHFAKQA:RESAGRETNGNOTVVRVF 180
|||||
QY 181 LLGKTPPEONHPDLSMDLKFESDKHODILMWNRYRDTFFNLSLKEVLFLEWVSTSCPDADF 240
|||||
Db 181 LLGKTPPEONHPDLSMDLKFESDKHODILMWNRYRDTFFNLSLKEVLFLEWVSTSCPDADF 240
|||||
QY 241 VFKGDDVFNTHHILNYLNSLSKSKAKDLFGDVIHNAHPHRRDKKLYIPEVFTGVY 300
|||||
Db 241 VFKGDDVFNTHHILNYLNSLSKSKAKDLFGDVIHNAHPHRRDKKLYIPEVFTGVY 300
|||||
QY 301 PPYAGGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKLGVLPEKHGKGFRTFDEE 360
|||||
Db 301 PPYAGGGGFLYSGPALLRLYSATSRVHLYPIDDVYTGMCLOKLGVLPEKHGKGFRTFDEE 360
|||||
QY 361 KKKKNCISYIDLMLVHSRKPQEMIDWISOLQSPNLKC 397
|||||
Db 361 KKKKNCISYIDLMLVHSRKPQEMIDWISOLQSPNLKC 397
|||||

RESULT 4
US-09-804-357-4
; Sequence 4, Application US/09804357
; Patent No. US20010024808A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/109001
; CURRENT APPLICATION NUMBER: US/09/804,357
```

; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/195,896
; PRIOR FILING DATE: 1998-11-19
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-804-357-4

Query Match 93.0%; Score 1979; DB 9; Length 365;
Best Local Similarity 100.0%; Pred. No. 2.9e-193;
Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 33 QDKNGKGGVVIIPKFKFKPPSTPRAYWNRQEKLNRYNPNILNRVANOTGELATSPNTSH 92
DB 1 QDKNGKGGVVIIPKFKFKPPSTPRAYWNRQEKLNRYNPNILNRVANOTGELATSPNTSH 60
QY 93 LSYCEPDSTVMTAVTDFNNLPDRFKDFLLYLRCRNYSLLDOPKKCAKPFLLLAIKSLI 152
DB 61 LSYCEPDSTVMTAVTDFNNLPDRFKDFLLYLRCRNYSLLDOPKKCAKPFLLLAIKSLI 120
QY 153 PHFARRQAIRESWGRETNGVQTVVRVFLGKTPPEDNHPDLSMDLKFESDKHODILMWN 212
DB 121 PHFARRQAIRESWGRETNGVQTVVRVFLGKTPPEDNHPDLSMDLKFESDKHODILMWN 180
QY 213 YRDTFFNLSLKEVLFLRWVSTSCPDAEVFVKGDODVFNTHHILNLYNSLSKSKAKDLFI 272
DB 181 YRDTFFNLSLKEVLFLRWVSTSCPDAEVFVKGDODVFNTHHILNLYNSLSKSKAKDLFI 240
QY 273 GDVIHNAHPHRDKKLYIPEVFTGYPPYAGGGFLYSGPALLRLLYSATSRVHLYPID 332
DB 241 GDVIHNAHPHRDKKLYIPEVFTGYPPYAGGGFLYSGPALLRLLYSATSRVHLYPID 300
QY 333 DVTGMCLOKLGVLPEKHGKGFRTDIEKNKKNICSVIDMLVHSRKPQEMIDIWSQLOS 392
DB 301 DVTGMCLOKLGVLPEKHGKGFRTDIEKNKKNICSVIDMLVHSRKPQEMIDIWSQLOS 360
QY 393 PNLKC 397
DB 361 PNLKC 365

RESULT 5
US-09-804-006-4
; Sequence 4, Application US/09804006
; Patent No. US20020119517A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/126001
; CURRENT APPLICATION NUMBER: US/09/804,006
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/292,228
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-804-006-4

Query Match 93.0%; Score 1979; DB 10; Length 365;
Best Local Similarity 100.0%; Pred. No. 2.9e-193;
Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 33 QDKNGKGGVVIIPKFKFKPPSTPRAYWNRQEKLNRYNPNILNRVANOTGELATSPNTSH 92
DB 1 QDKNGKGGVVIIPKFKFKPPSTPRAYWNRQEKLNRYNPNILNRVANOTGELATSPNTSH 60
QY 93 LSYCEPDSTVMTAVTDFNNLPDRFKDFLLYLRCRNYSLLDOPKKCAKPFLLLAIKSLI 152
DB 61 LSYCEPDSTVMTAVTDFNNLPDRFKDFLLYLRCRNYSLLDOPKKCAKPFLLLAIKSLI 120
QY 153 PHFARRQAIRESWGRETNGVQTVVRVFLGKTPPEDNHPDLSMDLKFESDKHODILMWN 212
DB 121 PHFARRQAIRESWGRETNGVQTVVRVFLGKTPPEDNHPDLSMDLKFESDKHODILMWN 180
QY 213 YRDTFFNLSLKEVLFLRWVSTSCPDAEVFVKGDODVFNTHHILNLYNSLSKSKAKDLFI 272
DB 181 YRDTFFNLSLKEVLFLRWVSTSCPDAEVFVKGDODVFNTHHILNLYNSLSKSKAKDLFI 240
QY 273 GDVIHNAHPHRDKKLYIPEVFTGYPPYAGGGFLYSGPALLRLLYSATSRVHLYPID 332
DB 241 GDVIHNAHPHRDKKLYIPEVFTGYPPYAGGGFLYSGPALLRLLYSATSRVHLYPID 300
QY 333 DVTGMCLOKLGVLPEKHGKGFRTDIEKNKKNICSVIDMLVHSRKPQEMIDIWSQLOS 392
DB 301 DVTGMCLOKLGVLPEKHGKGFRTDIEKNKKNICSVIDMLVHSRKPQEMIDIWSQLOS 360
QY 393 PNLKC 397
DB 361 PNLKC 365

RESULT 6
US-09-804-357-14
; Sequence 14, Application US/09804357
; Patent No. US20010024808A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/109001
; CURRENT APPLICATION NUMBER: US/09/804,357
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/195,896
; PRIOR FILING DATE: 1998-11-19
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-804-357-14

Query Match 87.5%; Score 1862; DB 9; Length 397;
Best Local Similarity 86.9%; Pred. No. 2.9e-181;
Matches 345; Conservative 19; Mismatches 33; Indels 0; Gaps 0;
QY 1 MSVGRRRVKKLIGILMMANVFYILVEVSKSSQDKNGKGGVVIIPKFKFWKPPSTPRAYWN 60
DB 1 MSVGRRIKKLIGILMMANVFYIFMEVSKSSQDKNGKGGVVIIPKFKFWKISTPPEAYWN 60
QY 61 REQEKLNRYNPNILNRVANOTGELATSPNTSHLSYCEPDSTVMTAVTDFNNLPDRFKDLFI 120
DB 61 REQEKLNRYNPNILNLTNQTGEAGRUSNLSHLSYCEPDSTVMTAVTDFNNLPDRFKDLFI 120
QY 121 LYLRCRNYSLLDOPKKCAKPFLLLAIKSLIPHARQAIRESWGRETNGVQTVVRVFL 180

Db 121 LYLRCRNYSLIDQPKCAKPFLLAIKSLTPHFARROAIRRESWGQESNAGNOTVVRVF 180
Qy 181 LLGKTPPEDNHPDLSMDLAFESDKHODILMWNRYRDTFFNLSLKEVLFWRWVSTSCPDDEF 240
Db 181 LLGKTPPEDNHPDLSMDLAFESDKHODILMWNRYRDTFFNLSLKEVLFWRWVSTSCPDDEF 240
Qy 241 VFKGDDDDVFVNTTHILNYLNSLSKSKAKOLFICGDIHINAGPHRDKKLYIPEVFTGYV 300
Db 241 VFKGDDDDVFVNTTHILNYLNSLSKSKAKOLFICGDIHINAGPHRDKKLYIPEVFTGYV 300
Qy 301 PPYAGGGFLYSGPALLRLYSATSRVHLYPIDDDVTGMCLOKGLVPEKHKGFRTFIDE 360
Db 301 PPYAGGGFLYSGPALLRLYSATSRVHLYPIDDDVTGMCLOKGLVPEKHKGFRTFIDE 360
Qy 361 KNKKNICSYIDLMLVHSRKPQEMIDWISQLOSAHLKC 397
Db 361 KNKKNICSYIDLMLVHSRKPQEMIDWISQLOSAHLKC 397
RESULT 7
US-09-972-912-2
: Sequence 2, Application US/09972912
: Patent No. US20020110867A1
: GENERAL INFORMATION:
: APPLICANT: SOPPET, DANIEL R.
: RUBEN, STEVEN M.
: TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
: STREET: 1100 NEW YORK AVENUE, SUITE 600
: CITY: WASHINGTON
: STATE: DC
: COUNTRY: US
: ZIP: 20005-3934
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION NUMBER: US/09/972,912
: FILING DATE: 10-Oct-2001
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 09/049,022
: FILING DATE: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: STEFFE, ERIC K.
: REGISTRATION NUMBER: 36,688
: REFERENCE/DOCKET NUMBER: 1488.0620001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 397 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-972-912-2
Query Match 87.5%; Score 1862; DB 10; Length 397;
Best Local Similarity 86.9%; Pred. No. 2,9e-181;
Matches 345; Conservative 19; Mismatches 33; Indels 0; Gaps 0;
Qy 1 MSVGRRRVKKLLGILMANVFIYIIVEVSKSSQDKNGKGVIIIPKEKFWKPESTPRAYWN 60
Db 1 MSVGRRRRIKLLGILMANVFIYIIMEVSKSSQDKNGKGVIIIPKEKFWKPESTPRAYWN 60
Qy 61 REQEKLNRYNPIILNRVANQTGELATSPNTSHLSYCEPDSTVMTAVTFDNNLPDRKDFL 120
Db 61 REQEKLNRYNPIILNRVANQTGELATSPNTSHLSYCEPDSTVMTAVTFDNNLPDRKDFL 120
Qy 121 LYLRCRNYSLIDQPKCAKPFLLAIKSLTPHFARROAIRRESWGQESNAGNOTVVRVF 180
Db 121 LYLRCRNYSLIDQPKCAKPFLLAIKSLTPHFARROAIRRESWGQESNAGNOTVVRVF 180
Qy 181 LLGKTPPEDNHPDLSMDLAFESDKHODILMWNRYRDTFFNLSLKEVLFWRWVSTSCPDDEF 240
Db 181 LLGKTPPEDNHPDLSMDLAFESDKHODILMWNRYRDTFFNLSLKEVLFWRWVSTSCPDDEF 240
Qy 241 VFKGDDDDVFVNTTHILNYLNSLSKSKAKOLFICGDIHINAGPHRDKKLYIPEVFTGYV 300
Db 241 VFKGDDDDVFVNTTHILNYLNSLSKSKAKOLFICGDIHINAGPHRDKKLYIPEVFTGYV 300

Db 61 REQEKLNRYNPIILNRVANQTGELATSPNTSHLSYCEPDSTVMTAVTFDNNLPDRKDFL 120
Qy 121 LYLRCRNYSLIDQPKCAKPFLLAIKSLTPHFARROAIRRESWGQESNAGNOTVVRVF 180
Db 121 LYLRCRNYSLIDQPKCAKPFLLAIKSLTPHFARROAIRRESWGQESNAGNOTVVRVF 180
Qy 181 LLGKTPPEDNHPDLSMDLAFESDKHODILMWNRYRDTFFNLSLKEVLFWRWVSTSCPDDEF 240
Db 181 LLGKTPPEDNHPDLSMDLAFESDKHODILMWNRYRDTFFNLSLKEVLFWRWVSTSCPDDEF 240
Qy 241 VFKGDDDDVFVNTTHILNYLNSLSKSKAKOLFICGDIHINAGPHRDKKLYIPEVFTGYV 300
Db 241 VFKGDDDDVFVNTTHILNYLNSLSKSKAKOLFICGDIHINAGPHRDKKLYIPEVFTGYV 300
RESULT 8
US-09-804-006-14
: Sequence 14, Application US/09804006
: Patent No. US20020119517A1
: GENERAL INFORMATION:
: APPLICANT: White, David
: APPLICANT: Zhou, Jiahangong
: TITLE OF INVENTION: LEPTIN INDUCED GENES
: FILE REFERENCE: 07334/126001
: CURRENT APPLICATION NUMBER: US/09/804,006
: CURRENT FILING DATE: 2001-03-12
: PRIOR APPLICATION NUMBER: US 09/292,228
: PRIOR FILING DATE: 1999-04-15
: PRIOR APPLICATION NUMBER: US 60/108,379
: PRIOR FILING DATE: 1998-10-29
: PRIOR APPLICATION NUMBER: US 09/150,857
: PRIOR FILING DATE: 1998-09-10
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 14
: LENGTH: 397
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-804-006-14
Query Match 87.5%; Score 1862; DB 10; Length 397;
Best Local Similarity 86.9%; Pred. No. 2,9e-181;
Matches 345; Conservative 19; Mismatches 33; Indels 0; Gaps 0;
Qy 1 MSVGRRRVKKLLGILMANVFIYIIVEVSKSSQDKNGKGVIIIPKEKFWKPESTPRAYWN 60
Db 1 MSVGRRRRIKLLGILMANVFIYIIMEVSKSSQDKNGKGVIIIPKEKFWKPESTPRAYWN 60
Qy 61 REQEKLNRYNPIILNRVANQTGELATSPNTSHLSYCEPDSTVMTAVTFDNNLPDRKDFL 120
Db 61 REQEKLNRYNPIILNRVANQTGELATSPNTSHLSYCEPDSTVMTAVTFDNNLPDRKDFL 120
Qy 121 LYLRCRNYSLIDQPKCAKPFLLAIKSLTPHFARROAIRRESWGQESNAGNOTVVRVF 180
Db 121 LYLRCRNYSLIDQPKCAKPFLLAIKSLTPHFARROAIRRESWGQESNAGNOTVVRVF 180
Qy 181 LLGKTPPEDNHPDLSMDLAFESDKHODILMWNRYRDTFFNLSLKEVLFWRWVSTSCPDDEF 240
Db 181 LLGKTPPEDNHPDLSMDLAFESDKHODILMWNRYRDTFFNLSLKEVLFWRWVSTSCPDDEF 240
Qy 241 VFKGDDDDVFVNTTHILNYLNSLSKSKAKOLFICGDIHINAGPHRDKKLYIPEVFTGYV 300
Db 241 VFKGDDDDVFVNTTHILNYLNSLSKSKAKOLFICGDIHINAGPHRDKKLYIPEVFTGYV 300

```

: TITLE OF INVENTION: LEPTIN INDUCED GENES
: FILE REFERENCE: 07334/109001
: CURRENT APPLICATION NUMBER: US/09/804,357
: CURRENT FILING DATE: 2001-03-12
: PRIOR APPLICATION NUMBER: US 09/195,896
: PRIOR FILING DATE: 1998-11-19
: PRIOR APPLICATION NUMBER: US 60/108,379
: PRIOR FILING DATE: 1998-10-29
: PRIOR APPLICATION NUMBER: US 09/150,857
: PRIOR FILING DATE: 1998-09-10
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 15
: LENGTH: 365
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-804-357-15

Query Match      81.08: Score 1725: DB 9: Length 365:
Best Local Similarity 86.88: Pred. No.2.4e-167:
Matches 317: Conservative 16: Mismatches 32: Indels 0: Gaps 0:

QY 33 QDKNGKGVIIIPKFKFWKPPSTPRAYWNREQELNRWYNPILNRVANOTGLATSPNTSH 92
   1:||||| ||||| : ||||| ||||| ||||| : ||||| |||||
Db 1 QEKNGKGVIIIPKFKFWK1KSTPEAYWNREQELNRQINPILSMLTNOTGAGRLSNLSH 60
   1:||||| ||||| : ||||| ||||| ||||| : ||||| |||||
QY 93 LSYCEPDSTVMTAVTDENNLPDRKDFLYLRCRNYSLIIDQPKCAKKPFLLAIAKSLI 152
   1:||||| | : ||||| ||||| ||||| ||||| ||||| |||||
Db 61 LNYCEPDURVTSVYTGFPNNLPDRKDFLYLRCRNYSLIIDQPKCAKKPFLLAIAKSLT 120
   1:||||| ||||| : ||||| ||||| ||||| ||||| ||||| |||||
QY 153 PHEARROAIRESWGREFNMGVNOTWVRFVLGLGKTPPEDNHPDLSDMLKFESEKHODILMN 212

```

```

Db      121 PHFAKRAQIREWGQESNAGNQTVVRVILGQTTPEDNHDPDLSOMLKFESKHKQDILWNN 180
QY      213 YRDTFFNLSEKVLFLRWVSTSCPDAEVFKGDDDVFNTHHILNYLNSLSKSAKDLFI 272
Db      181 YRDTFFNLSEKVLFLRWVSTSCPDTEFVFKGDDDVFNTHHILNYLNSLSKSAKDLFI 240
QY      273 GDVTHNACPHRDKKLYIPEVYFTGVYPPYAGGGGFLYSGPALLRLYSATSRVHLYPID 332
Db      241 GDVTHNACPHRDKKLYIPEVYSGLYPPYAGGGGFLYSGHLALRLYHITDQVHLYPID 300
QY      333 DVTYGMCLQKLGVLPEKHKGFRTFDTEENKKNICSYIDLMLVHSRKPOEMIDIWSOLQS 392
Db      301 DVTYGMCLQKLGVLPEKHKGFRTFDTEENKKNICSYVDLMLVHSRKPOEMIDIWSOLQS 360
QY      393 PNLC 397
Db      361 AHLK 365

RESULT 11
US-09-804-006-15
; Sequence 15, Application US/09804006
; Patent NO. US20020119517A1
; GENERAL INFORMATION:
; APPLICANT: white, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/126001
; CURRENT APPLICATION NUMBER: US/09/804,006
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/292,228
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 365

```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-804-006-15

Query Match      81.0%; Score 1725; DB 10; Length 365;
Best Local Similarity 86.8%; Pred. No. 2.4e-167;
Matches 317; Conservative 16; Mismatches 32; Indels 0; Gaps 0;

Qy 33 QDRNGKGGVLIPEKFKWPPSTPRAYWNRQEKLNRYNFIILNRVANOTGELATSPNTSH 92
Db 1 QEKNGKEVLIPEKFKWISTPPEAYWNRQEKLNRYNFIILNMQTQAGRLSNISH 60

Qy 93 LSCPEPSTYMTAVTFNNLPDRFKDFLLYLRCRNTSLIDOPKKCAKPFLLLAIKSLI 152
Db 61 LNYCEPDLRYTSVTVFNNLPDRFKDFLLYLRCRNTSLIDOPKKCAKPFLLLAIKSLI 120

Qy 153 PHFARQAIRESWGRETNGVQTVVRFVLLGKTPPEDNHPDLSMDLKFESKHQDILMWN 212
Db 121 PHFARQAIRESWGQESNAGNQTVVRFVLLGQTPPEDNHPDLSMDLKFESKHQDILMWN 180

Qy 213 YRDTFNLSLKEVLFRLRWSTSCPDAEFVKGDGDDVFVNTHHILNYLSLSKSKAKDLFI 272
Db 181 YRDTFNLSLKEVLFRLRWSTSCPDTEFVKGDGDDVFVNTHHILNYLSLSKSKAKDLFI 240

Qy 273 GDVIHAGPHRDKKLYIPEVYTGVPYAGGGFLYSGPALLRLYSATSRVHLYPID 332
Db 241 GDVIHAGPHRDKKLYIPEVYSGLYPYAGGGFLYSGHLALRLYHITDQVHLYPID 300

Qy 333 DVTGMCLOKLGVLPEKHGKFRFTDIEKNKKNICSYIDLMLVHSRKPQEMIDINWSQLOS 392
Db 301 DVTGMCLOKLGVLPEKHGKFRFTDIEKNKKNICSYVDLMLVHSRKPQEMIDINWSQLOS 360

Qy 393 PNLC 397
Db 361 AHLK 365

RESULT 12
US-09-739-451-4
; Sequence 4, Application US/09739451
; Patent No. US20010024813A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Dendriac and Brainiac-3
; FILE REFERENCE: PF464
; CURRENT APPLICATION NUMBER: US/09/739,451
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/213,364
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: 60/077,687
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/108,928
; PRIOR FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-739-451-4

Query Match      30.8%; Score 656; DB 9; Length 352;
Best Local Similarity 40.2%; Pred. No. 2.9e-58;
Matches 146; Conservative 53; Mismatches 122; Indels 42; Gaps 8;

Qy 40 GVIIPEKFKWPPSTPRAYWNRQEKLNRYNFIILNRVANOTGELATSPNTSHLSYCEPD 99
Db 19 GCLLFRLKKAAPETPR-----PTSLSGAPPTPRHSRCPN 54

Qy 100 STYMTAVTFNNLPDRFKDFLLYLRCRNTSLIDOPKKCAKPFLLLAIKSLIPHFARQ 159
Db 55 HTVSSASL---SLPSRHLFLTYRHCRNFSILL-EPGSGSKDTFLLLAIKSQPGHVERA 110

Qy 160 AIRESWGRETNGVQTVVRFVLL---CKTPPEDNHPDLSMDLKFESKHQDILMWNRYDT 216
Db 111 AIRSTWGRWGDLGPGALKVFLLGVAGSAPP-----AQLLAYESREFDDILQWDFTE 163

Qy 217 FFNLSLKEVLFRLRWSTSCPDAEFVKGDGDDVFVNTHHILNYLSLSKSKAKDLFI 276
Db 164 FFNLTJLKLHLQWVAACPAHFMLKGGDDVFVHPVNVLEFLDGM--DPAQDQLLVGDI 221

Qy 277 HNAGPHRDKKLYIPEVYTGVPYAGGGFLYSGPALLRLYSATSRVHLYPID 335
Db 222 QOALPNENTKVKYFIPPSMYRATHYPPYAGGGVYMSRATVRRLQAIMEDAELFIDV 281

Qy 336 TGMCLQKLGVLPEKHGKFRFTDIEEK-NKKNICSYIDLMLVHSRKPQEMIDINWSQLOS 394
Db 282 VGMCLRLGLSPMHAGFKFTGIRRLDPLDPCLYRGLLLVHRLSPLEMTWMLVTDG 341

Query Match      30.8%; Score 656; DB 9; Length 353;
Best Local Similarity 40.2%; Pred. No. 3e-58;
Matches 146; Conservative 53; Mismatches 122; Indels 42; Gaps 8;

Qy 40 GVIIPEKFKWPPSTPRAYWNRQEKLNRYNFIILNRVANOTGELATSPNTSHLSYCEPD 99
Db 19 GCLLFRLKKAAPETPR-----PTSLSGAPPTPRHSRCPN 54

Qy 100 STYMTAVTFNNLPDRFKDFLLYLRCRNTSLIDOPKKCAKPFLLLAIKSLIPHFARQ 159
Db 55 HTVSSASL---SLPSRHLFLTYRHCRNFSILL-EPGSGSKDTFLLLAIKSQPGHVERA 110

Qy 160 AIRESWGRETNGVQTVVRFVLL---CKTPPEDNHPDLSMDLKFESKHQDILMWNRYDT 216
Db 111 AIRSTWGRWGDLGPGALKVFLLGVAGSAPP-----AQLLAYESREFDDILQWDFTE 163

Qy 217 FFNLSLKEVLFRLRWSTSCPDAEFVKGDGDDVFVNTHHILNYLSLSKSKAKDLFI 276
Db 164 FFNLTJLKLHLQWVAACPAHFMLKGGDDVFVHPVNVLEFLDGM--DPAQDQLLVGDI 221

Qy 277 HNAGPHRDKKLYIPEVYTGVPYAGGGFLYSGPALLRLYSATSRVHLYPID 335
Db 222 QOALPNENTKVKYFIPPSMYRATHYPPYAGGGVYMSRATVRRLQAIMEDAELFIDV 281

Qy 336 TGMCLQKLGVLPEKHGKFRFTDIEEK-NKKNICSYIDLMLVHSRKPQEMIDINWSQLOS 394
Db 282 VGMCLRLGLSPMHAGFKFTGIRRLDPLDPCLYRGLLLVHRLSPLEMTWMLVTDG 341
```

```
Qy 160 AIRESWGRETNGVQTVVRFVLL---CKTPPEDNHPDLSMDLKFESKHQDILMWNRYDT 216
Db 111 AIRSTWGRWGDLGPGALKVFLLGVAGSAPP-----AQLLAYESREFDDILQWDFTE 163

Qy 217 FFNLSLKEVLFRLRWSTSCPDAEFVKGDGDDVFVNTHHILNYLSLSKSKAKDLFI 276
Db 164 FFNLTJLKLHLQWVAACPAHFMLKGGDDVFVHPVNVLEFLDGM--DPAQDQLLVGDI 221

Qy 277 HNAGPHRDKKLYIPEVYTGVPYAGGGFLYSGPALLRLYSATSRVHLYPID 335
Db 222 QOALPNENTKVKYFIPPSMYRATHYPPYAGGGVYMSRATVRRLQAIMEDAELFIDV 281

Qy 336 TGMCLQKLGVLPEKHGKFRFTDIEEK-NKKNICSYIDLMLVHSRKPQEMIDINWSQLOS 394
Db 282 VGMCLRLGLSPMHAGFKFTGIRRLDPLDPCLYRGLLLVHRLSPLEMTWMLVTDG 341

Qy 395 LK 397
Db 342 LK 344

RESULT 13
US-09-739-451-12
; Sequence 12, Application US/09739451
; Patent No. US20010024813A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Dendriac and Brainiac-3
; FILE REFERENCE: PF464
; CURRENT APPLICATION NUMBER: US/09/739,451
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/213,364
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: 60/077,687
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/108,928
; PRIOR FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-739-451-12

Query Match      30.8%; Score 656; DB 9; Length 353;
Best Local Similarity 40.2%; Pred. No. 3e-58;
Matches 146; Conservative 53; Mismatches 122; Indels 42; Gaps 8;

Qy 40 GVIIPEKFKWPPSTPRAYWNRQEKLNRYNFIILNRVANOTGELATSPNTSHLSYCEPD 99
Db 19 GCLLFRLKKAAPETPR-----PTSLSGAPPTPRHSRCPN 54

Qy 100 STYMTAVTFNNLPDRFKDFLLYLRCRNTSLIDOPKKCAKPFLLLAIKSLIPHFARQ 159
Db 55 HTVSSASL---SLPSRHLFLTYRHCRNFSILL-EPGSGSKDTFLLLAIKSQPGHVERA 110

Qy 160 AIRESWGRETNGVQTVVRFVLL---CKTPPEDNHPDLSMDLKFESKHQDILMWNRYDT 216
Db 111 AIRSTWGRWGDLGPGALKVFLLGVAGSAPP-----AQLLAYESREFDDILQWDFTE 163

Qy 217 FFNLSLKEVLFRLRWSTSCPDAEFVKGDGDDVFVNTHHILNYLSLSKSKAKDLFI 276
Db 164 FFNLTJLKLHLQWVAACPAHFMLKGGDDVFVHPVNVLEFLDGM--DPAQDQLLVGDI 221

Qy 277 HNAGPHRDKKLYIPEVYTGVPYAGGGFLYSGPALLRLYSATSRVHLYPID 335
Db 222 QOALPNENTKVKYFIPPSMYRATHYPPYAGGGVYMSRATVRRLQAIMEDAELFIDV 281

Qy 336 TGMCLQKLGVLPEKHGKFRFTDIEEK-NKKNICSYIDLMLVHSRKPQEMIDINWSQLOS 394
Db 282 VGMCLRLGLSPMHAGFKFTGIRRLDPLDPCLYRGLLLVHRLSPLEMTWMLVTDG 341
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2003, 13:13:22 ; Search time 4471.3 Seconds
(without alignments)
10942.656 Million cell: updates/sec

Title: US-09-804-357B-1
Perfect score: 1196
Sequence: 1 agatgagctg9gggctcga.....gtcttaattaaatgctga il96

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.:

- 1: gb_ba.:
- 2: gb_htg.:
- 3: gb_in.:
- 4: gb_ov.:
- 5: gb_ov.:
- 6: gb_pat.:
- 7: gb_ph.:
- 8: gb_pl.:
- 9: gb_pr.:
- 10: gb_ro.:
- 11: gb_sts.:
- 12: gb_sy.:
- 13: gb_un.:
- 14: gb_vi.:
- 15: em_ba.:
- 16: em_fun.:
- 17: em_hum.:
- 18: em_in.:
- 19: em_mu.:
- 20: em_mu.:
- 21: em_or.:
- 22: em_ov.:
- 23: em_pat.:
- 24: em_ph.:
- 25: em_pl.:
- 26: em_ro.:
- 27: em_sts.:
- 28: em_un.:
- 29: em_vi.:
- 30: em_htg_hum.:
- 31: em_htg_inv.:
- 32: em_htg_other.:
- 33: em_htg_mus.:
- 34: em_htg_pln.:
- 35: em_htg_rod.:
- 36: em_htg_mam.:
- 37: em_htg_vrt.:
- 38: em_sy.:
- 39: em_htgo_hum.:
- 40: em_htgo_mus.:
- 41: em_htgo_other.:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1196	100.0	2468	10	BC009075	BC009075 Mus muscu
2	1196	100.0	2672	10	AY043479	AY043479 Mus muscu
3	1196	100.0	210863	10	AL772364	AL772364 Mouse DNA
4	1192.8	99.7	1260	10	AF092050	AF092050 Mus muscu
5	1064.4	89.0	178009	2	AC119549	AC119549 Rattus no
6	1056.4	88.3	318862	2	AC109547	AC109547 Rattus no
7	906.4	75.8	1831	6	AF092051	AF092051 Homo sapi
8	906.4	75.8	1912	6	BD093191	BD093191 Useful po
9	906.4	75.8	1912	9	AB049584	AB049584 Homo sapi
10	906.4	75.8	2454	9	BC030579	BC030579 Homo sapi
11	906.4	75.8	2714	9	AF288208	AF288208 Homo sapi
12	906.4	75.8	2811	9	BC047933	BC047933 Homo sapi
13	906.4	75.8	99088	9	AC093401	AC093401 Homo sapi
14	893	74.7	1197	9	HSA6077	AJ006077 Homo sapi
15	885.4	74.0	2516	9	AF288209	AF288209 Homo sapi
16	848.6	71.0	2442	6	BD156964	BD156964 Primer fo
17	848.6	71.0	2442	9	AK002009	AK002009 Homo sapi
18	724.8	60.6	126480	2	AC040954	AC040954 Mus muscu
19	713.8	59.7	172509	2	AC024716	AC024716 Homo sapi
20	713.8	59.7	200191	9	AL390027	AL390027 Human DNA
21	713.8	59.7	288669	2	AL390297	AL390297 Homo sapi
22	521.8	43.6	832	6	BD149792	BD149792 Primer fo
23	415.4	34.7	170370	2	BX469915	BX469915 Danio rer
24	415	34.7	2162	5	AF321831	AF321831 Danio rer
25	375	31.4	1682	5	AF321828	AF321828 Danio rer
26	370	30.9	556	6	BD071997	BD071997 Secreted
27	362.6	30.3	197598	2	AC144887	AC144887 Bos tauru
28	203.8	17.0	1245	6	AX513603	AX513603 Sequence
29	186.2	15.6	2260	5	AF321827	AF321827 Danio rer
30	185.6	15.5	1564	5	AF321829	AF321829 Danio rer
31	178	14.9	209192	2	AC102609	AC102609 Mus muscu
32	174.8	14.6	1502	10	AF502429	AF502429 Mus muscu
33	174.8	14.6	2367	10	BC031187	BC031187 Mus muscu
34	173.2	14.5	230799	2	AC097668	AC097668 Rattus no
35	172.6	14.4	1434	9	AF502430	AF502430 Homo sapi
36	172.6	14.4	168880	9	AC017104	AC017104 Homo sapi
37	171	14.3	1360	6	AX451423	AX451423 Sequence
38	169.8	14.2	172978	2	BX530031	BX530031 Danio rer
39	169.8	14.2	233276	2	BX511175	BX511175 Danio rer
40	169.4	14.2	1134	6	AR202302	AR202302 Sequence
41	163	13.6	1191	6	AR217803	AR217803 Sequence
42	160	13.4	2003	6	AX513601	AX513601 Sequence
43	157.2	13.1	194363	5	AL954634	AL954634 Zebrafish
44	157.2	13.1	206683	2	BX323592	BX323592 Danio rer
45	156.8	13.1	1167	6	AR217813	AR217813 Sequence

ALIGNMENTS

RESULT 1
BC009075 2468 bp mRNA linear ROD 16-APR-2003
LOCUS BC009075
DEFINITION Mus musculus UDP-GlcNAc:betaGal
beta-1,3-N-acetylglucosaminyltransferase 1, mRNA (cDNA clone
MGC:6892 IMAGE:2654354), complete cds.
ACCESSION BC009075
VERSION BC009075.1 GI:14290591
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2468)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,

Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D., Altschul, S.F., Zebberg, S., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.C., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 2468)

Strausberg, R.

Direct Submission

Submitted (30-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NTI-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabbs@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Sequencing Center

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loussouad, H., Kowis, C.R., Speed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAP Plate: 5 Row: e Column: 16

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 9938023, location/Qualifiers

FEATURES

source

1..2468

/organism="Mus musculus"

/mol_type="mRNA"

/strain="NWRI"

/db_xref="taxon:10090"

/clone="MGC:6892 IMAGE:2654354"

/tissue_type="Mammary tumor. WAP-Tag model. 5 months old, gross tissue."

/clone_lib="NCI_CGAP_Mam4"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

1..2468

/gene="B3gnt1"

/note="synonyms: Beta3gnt, beta-3gnt, B3Galt6"

/db_xref="LocusID:53625"

/db_xref="MGI:1889505"

129..1322

/codon_start=1

/product="UDP-GlcNAc:betaGai

beta-1,3-N-acetylglucosaminyltransferase 1"

/protein_id="AAH09075.1"

/db_xref="GI:14290592"

gene

CDS

/db_xref="LocusID:53625"

/translation="MSVGRVRVLLGLILMANVFYILVEVSKNSQDNRKGGVILP
KEKFWKPPSPRAYWNRQEKLNRYNPLNRVANQOTGLATSPNTSHLSYEPDSTV
MTAVTFDNPLPDRFKDFLLVLCRNTSLIDQPKKAKKPFLLATKSLIPHFARQA
IRSGRETNGVNOTVVRVELLCKTPEDNHPDLSDMLKFECDKHODILMNYRDFEF
NLSLKEVLPFLRWYSTCPDAEFYVFGDDDDVFYVTHLLNLSLSKSKAKDLFIGDVI
HNAGPHDKDKYIIPFYVPPYAGGFLYSGPLALRSLYSRVHLYPIDDV
YTMCKLQKGLVPEKHGKFTFDIEKNKNKICSYIDLMLVHRSKQEMIDWSQLQS
PNLKC"

BASE COUNT 685 a 509 c 564 g 710 t

ORIGIN

Query Match 100.0%; Score 1196; DB 10; Length 2468;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGATGAGTGTGGGGCGTCGAAGAGTCAAGTTGCTGGGCATCCTGATGATGGCAATGTCT 60
DB 127 AGATGAGTGTGGGGCGTCGAAGAGTCAAGTTGCTGGGCATCCTGATGATGGCAATGTCT 186

QY 61 TCATTTATTTGATGTGGAAAGTCTCCAAAACAGTAGCCCAAGACAAAATGGAAAGGAG 120
DB 187 TCATTTATTTGATGTGGAAAGTCTCCAAAACAGTAGCCCAAGACAAAATGGAAAGGAG 246

QY 121 GAGTAATAATCCCGAAAGAGAGTTCCTGAAGCCACCAGCAGCTCCCGGGCAGTACTGGA 180
DB 247 GAGTAATAATCCCGAAAGAGAGTTCCTGAAGCCACCAGCAGCTCCCGGGCAGTACTGGA 306

QY 181 ACAGGGACAGGAGAGCTGAACAGTGTGTACATCCCATCTTTGAACAGGGTGGCCAAATC 240
DB 307 ACAGGGACAGGAGAGCTGAACAGTGTGTACATCCCATCTTTGAACAGGGTGGCCAAATC 366

QY 241 AGACAGGGAGTACCCACATCTCCAAACACAGTACCTGAGCTATTGTGAACACAGACT 300
DB 367 AGACAGGGAGTACCCACATCTCCAAACACAGTACCTGAGCTATTGTGAACACAGACT 426

QY 301 CGAGGGTCATGACAGCTGTGACAGATTTTAAATCTGCCGGACAGATTTTAAAGACTTTC 360
DB 427 CGAGGGTCATGACAGCTGTGACAGATTTTAAATCTGCCGGACAGATTTTAAAGACTTTC 486

QY 361 TCTTGTATTTGAGATGCGGAATTTACTCGCTGCTTATAGATCAACGACAAATGTGCA 420
DB 487 TCTTGTATTTGAGATGCGGAATTTACTCGCTGCTTATAGATCAACGACAAATGTGCA 546

QY 421 AGAAGCCCTTCTTACTATTGGCGATAAGTCCCTCATTTCCACATTTTCCGAGAGCAAG 480
DB 547 AGAAGCCCTTCTTACTATTGGCGATAAGTCCCTCATTTCCACATTTTCCGAGAGCAAG 606

QY 481 CAATTCGGAGTCTTGGGGCGGAGAACCAACAGTAGGGAACACAGATAGTAGGTCT 540
DB 607 CAATTCGGAGTCTTGGGGCGGAGAACCAACAGTAGGGAACACAGATAGTAGGTCT 666

QY 541 TCCTGTTGGGCAAGACACCCCGAGGACCAACACCTGACCTTTCCGACATCTTAAAGT 600
DB 667 TCCTGTTGGGCAAGACACCCCGAGGACCAACACCTGACCTTTCCGACATCTTAAAGT 726

QY 601 TTGAGAGTGACAAGCACACAGGACATCCTCATGTGTGAACTATAGAGACACATTTCAACC 660
DB 727 TTGAGAGTGACAAGCACACAGGACATCCTCATGTGTGAACTATAGAGACACATTTCAACC 786

QY 661 TGTCCCTGAAGGAAGTGTCTTCTTAGTGGGTGAGCAGCTTCTGTCAGAGCGCAGAGT 720
DB 787 TGTCCCTGAAGGAAGTGTCTTCTTAGTGGGTGAGCAGCTTCTGTCAGAGCGCAGAGT 846

QY 721 TTGCTTCAAGGGCGATGATGACGCTGTTGTGAACACCCATCATCTTAAATTAAGTGA 780
DB 847 TTGCTTCAAGGGCGATGATGACGCTGTTGTGAACACCCATCATCTTAAATTAAGTGA 906

QY 781 ATAGCTTATCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 840
DB 907 ATAGCTTATCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 966

QY 841 GGCCTCACCGGGATGAAGAACTGAAGTACTACATCCCAAGAGTCTTCTTACACCGGGCTCT 900

[illegible]

QY 1021 TTCAGAACTGGGCTTGTTCAGAGAGACAGAAAGCCTTCAGACATTTGATTAATGAAG 1080
|||||
Db 1228 TTCAGAACTGGGCTTGTTCAGAGAGACAGAAAGCCTTCAGACATTTGATTAATGAAG 1287
|||||
QY 1081 AGAAAAATAGAAAAATATTTGCTTATATAGACCTTAATGTTAGTACATAGCAGAAAC 1140
|||||
Db 1288 AGAAAAATAGAAAAATATTTGCTTATATAGACCTTAATGTTAGTACATAGCAGAAAC 1347
|||||
QY 1141 CTCAGAGATGATGATATCTGGTCTCAGCTGCAAGTCTCAATTTAAATGCTGA 1196
|||||
Db 1348 CTCAGAGATGATGATATCTGGTCTCAGCTGCAAGTCTCAATTTAAATGCTGA 1403
|||||

RESULT 3

AL772364/c 210863 bp DNA linear ROD 18-FEB-2003
LOCUS Mouse DNA sequence from clone RP23-242C19 on chromosome 11,
DEFINITION complete sequence.

AL772364

ACCESSION

AL772364.20 GI:28412567

KEYWORDS

HTG.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 210863)

Andrew.R.

Direct Submission

Submitted (18-FEB-2003) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk

On Feb 18, 2003 this sequence version replaced gi:28300672.

Sequence from the Mouse Genome Sequencing Consortium whole genome

shotgun may have been used to confirm this sequence. Sequence data

from the whole genome shotgun alone has only been used where it has

a phred quality of at least 30.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest, except on the rare

occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession:

numbers given in the feature table with their source databases:

Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep

from the RPI-23 Mouse BAC Library

constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6.

Location/Qualifiers

1. .210863

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/Chromosome="11"

/clone="RP23-242C19"

/clone_lib="RPI-23"

FEATURES

source

BASE COUNT	62433 a	47427 c	45742 g	55261 t	ORIGIN
Query Match	100.0%	Score 1196;	DB 10;	Length 210863;	
Best Local Similarity	100.0%;	Pred. No. 0;			
Matches 1196;	Conservative	0;	Mismatches	0;	Indels
0;	Gaps	0;			
QY 1	AGATGAGTGTGGGGTTCGAAGAGTCAAGTTGCTGGGCATCTCTGATGATGGCAATGTCT	60			
Db 71238	AGATGAGTGTGGGGTTCGAAGAGTCAAGTTGCTGGGCATCTCTGATGATGGCAATGTCT	71179			
QY 61	TCATTTATTTGATTTGCGAAGTCTCCAAAGACAGTACCGACAGCAAAATGGAAGGAG	120			
Db 7-178	TCATTTATTTGATTTGCGAAGTCTCCAAAGACAGTACCGACAGCAAAATGGAAGGAG	71119			
QY 121	GAGTAATATCCGAAAGAGAAAGTTCTGGAAGCCACCCAGCACCTCCCGGGCATACTGGA	180			
Db 71118	GAGTAATATCCGAAAGAGAAAGTTCTGGAAGCCACCCAGCACCTCCCGGGCATACTGGA	71059			
QY 181	ACAGGGAACAGAGAGAGTGAACAGGTGGTACATCCATCTTGAACAGGGTGGCCAAATC	240			
Db 71058	ACAGGGAACAGAGAGAGTGAACAGGTGGTACATCCATCTTGAACAGGGTGGCCAAATC	70999			
QY 241	AGACAGGGGAGCTAGCCACATCTCCAAACACAGTCACTGAGCTATTGTGAACAGACT	300			
Db 70998	AGACAGGGGAGCTAGCCACATCTCCAAACACAGTCACTGAGCTATTGTGAACAGACT	70939			
QY 301	CGACGGTCTATGACAGCTGTGACAGATTTTAATAATCTGCCGGACAGATTTAAAGACTTC	360			
Db 70938	CGACGGTCTATGACAGCTGTGACAGATTTTAATAATCTGCCGGACAGATTTAAAGACTTC	70879			
QY 361	TCTTGTATTTGAGATGCGGAAATCTACGCTGCTTATAGATCAACCGAAGAAATGTGCA	420			
Db 70878	TCTTGTATTTGAGATGCGGAAATCTACGCTGCTTATAGATCAACCGAAGAAATGTGCA	70819			
QY 421	AGAAGCCCTTCTTACTATTGGCGATTAAGTCCCTCATTCACATTTTGCACGAAGGCAAG	480			
Db 70818	AGAAGCCCTTCTTACTATTGGCGATTAAGTCCCTCATTCACATTTTGCACGAAGGCAAG	70759			
QY 481	CAATTTCGGGAGTCTTGGGGCGGAGAACCAACAGTAGGGAACACACAGTAGTGGGTCT	540			
Db 70758	CAATTTCGGGAGTCTTGGGGCGGAGAACCAACAGTAGGGAACACACAGTAGTGGGTCT	70699			
QY 541	TCCTGTTGGGAAGACACCCCGGAGGACAAACACCTGACCTTCGAGCATGCTTAAGT	600			
Db 70698	TCCTGTTGGGAAGACACCCCGGAGGACAAACACCTGACCTTCGAGCATGCTTAAGT	70639			
QY 601	TTGAGAGTGACAAGCACACAGACATCCTCATGTGGAACATATAGACACATTTCTCAACC	660			
Db 70638	TTGAGAGTGACAAGCACACAGACATCCTCATGTGGAACATATAGACACATTTCTCAACC	70579			
QY 661	TGTCCTTGAAGGAAGTGTGTTTCTTAGTGGGTGAGCAGTTCCTGTCACAGACGACAGT	720			
Db 70578	TGTCCTTGAAGGAAGTGTGTTTCTTAGTGGGTGAGCAGTTCCTGTCACAGACGACAGT	70519			
QY 721	TTGCTTCAAGGGCGATGACAGTGTGTTGTAACACCCATCATCTTAAATTAATGTA	780			
Db 705-8	TGTCCTTCAAGGGCGATGACAGTGTGTTGTAACACCCATCATCTTAAATTAATGTA	70459			
QY 781	ATAGCTTATCCAAGAGCAAGCCCAAGACTTGTTCATAGGTGAGTGCATCCACAATGCTG	840			
Db 70458	ATAGCTTATCCAAGAGCAAGCCCAAGACTTGTTCATAGGTGAGTGCATCCACAATGCTG	70399			
QY 841	GSCCTCACCGGATTAAGAACTGAAGTACTACATCCAGAGTCTTCTACACCGGCTCT	900			
Db 70398	GSCCTCACCGGATTAAGAACTGAAGTACTACATCCAGAGTCTTCTACACCGGCTCT	70339			
QY 901	ACCCACCGTATCCCGGGGTGGTGGATTCCTGTACTCCGGCCCCCTTGCCTTGAGGCTGT	960			
Db 70338	ACCCACCGTATCCCGGGGTGGTGGATTCCTGTACTCCGGCCCCCTTGCCTTGAGGCTGT	70279			
QY 961	ACAGTGCACACTAGCCGGGTCCATCTTACCCCTATTGATGATGTTTATACGGGAATGTGCC	1020			

Db	70278	ACAGIGGAGCTAGCGGGGCCATCTCTACCTATTGATGATGTTTATACGGGAATGACC	70219
Qy	1021	TTGAGAACTGGGCGCTTGTTCAGAGAGACAAAGGCTTCAGGACATTTGATATTGAAG	1080
Db	70218	TTGAGAACTGGGCGCTTGTTCAGAGAGACAAAGGCTTCAGGACATTTGATATTGAAG	70159
Qy	1081	AGAAATATAGAAATATTTGTTCTCTATATAGACCTAATGTTAGTACATAGCAGAAAC	1140
Db	70158	AGAAATATAGAAATATTTGTTCTCTATATAGACCTAATGTTAGTACATAGCAGAAAC	70099
Qy	1141	CTCAAGAGATGATGATATCTGCTCTCAGTTCGCAAGTCTAATTTAAATGCTGA	1196
Db	70098	CTCAAGAGATGATGATATCTGCTCTCAGTTCGCAAGTCTAATTTAAATGCTGA	70043
RESULT 4			
AF092050		1260 bp mRNA linear ROD 09-AUG-2000	
LOCUS		Mus musculus beta-1,3-N-acetylglucosaminyltransferase mRNA,	
DEFINITION		complete cds.	
ACCESSION		AF092050	
VERSION		AF092050.2 GI:9755418	
KEYWORDS			
SOURCE		Mus musculus (house mouse)	
ORGANISM			
REFERENCE			
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
TITLE		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
JOURNAL		1. (bases 1 to 1260)	
MEDLINE		Zhou,D., Dinter,A., Gutierrez Gallego,R., Kamberling,J.P.,	
PUBMED		Vilegandhant,J.F., Berger,E.G. and Hennet,T.	
REFERENCE		A beta-1,3-N-acetylglucosaminyltransferase with	
AUTHORS		poly-N-acetylglucosamine synthase activity is structurally related	
TITLE		to beta-1,3-galactosyltransferases	
JOURNAL		Proc. Natl. Acad. Sci. U.S.A. 96 (2), 406-411 (1999)	
MEDLINE		99110903	
PUBMED		9892646	
REFERENCE		2. (bases 1 to 1260)	
AUTHORS		Zhou,D., Berger,E.G. and Hennet,T.	
TITLE		Direct Submission	
JOURNAL		Submitted (15-SEP-1998) Physiology, University of Zurich,	
REFERENCE		Winterthurerstrasse 190, Zurich 8057, Switzerland	
AUTHORS		3. (bases 1 to 1260)	
TITLE		Zhou,D., Berger,E.G. and Hennet,T.	
JOURNAL		Submitted (09-AUG-2000) Physiology, University of Zurich,	
REFERENCE		Winterthurerstrasse 190, Zurich 8057, Switzerland	
AUTHORS		Sequence update by submitter	
TITLE		On Aug 9, 2000 this sequence version replaced gi:4191391.	
JOURNAL		Location/Qualifiers	
REMARK		1. .1260	
FEATURES		/organism="Mus musculus"	
SOURCE		/mol_type="mRNA"	
		/strain="ICR outbred"	
		/db_xref="taxon:10090"	
		/tissue_type="brain"	
		/dev_stage="newborn"	
		30. .1223	
		/EC_number="2.4.1.149"	
		/note="glycosyltransferase; b3GnT"	
		/codon_start=1	
		/product="beta-1,3-N-acetylglucosaminyltransferase"	
		/protein_id="AAB09763.2"	
		/db_xref="GI:9755419"	
		/translation="MSVRRRVKLLGLIMANVFYLYIVEVSKNSQDKNGGVIIIP	
		KERFAPSPRAYRREKLRNYPNLRVANOTGELATSPNTHSLSYCEPDSTV	
		IRTSWRETNVGNQTVRVFLGKTPPEDNPDLSMLKFESDKHODILMWNRYRTFF	
		NLSLKEVLFRLWVSTCPDAEFVKGDVDVFNTHILNLSLSKSKAKDLF IGDI	
		HNAGPHSDKKLYI IPEVFTGVYPPYAGGGFLYSGPLALRLYSATSRVHLXPIDDV	
		YTCMLQKLGVLPEKHGRTFIDIEKNKNKICSYIDLMLVHSRFPQEMIDISWLQS	
		PNLKC.	
BASE COUNT		370 a 282 c 307 g 301 t	
ORIGIN			

Qy	1	AGATCAGTGTGGGGTTCGAAGAGTCAAGTTGCTGGGCATCTCTGATGATGCAAAATGCT	60
Db	28	AGATCAGTGTGGGGTTCGAAGAGTCAAGTTGCTGGGCATCTCTGATGATGCAAAATGCT	87
Qy	61	TCATTATTATTGATTGTGGAAGTCTCCAAAACAGTAGCAAGACAAAATGGAAGGAG	120
Db	88	TCATTATTATTGATTGTGGAAGTCTCCAAAACAGTAGCAAGACAAAATGGAAGGAG	147
Qy	121	GAGTAAATATCCGAAAGAGAGTTCTGGAAGCCACCAGCAGCTCCCGGGCATACTGGA	180
Db	148	GAGTAAATATCCGAAAGAGAGTTCTGGAAGCCACCAGCAGCTCCCGGGCATACTGGA	207
Qy	181	ACAGGGAACAGAGAGCTGAACAGTGTCAATCCCATCTTGAACAGGTGGCCAAATC	240
Db	208	ACAGGGAACAGAGAGCTGAACAGTGTCAATCCCATCTTGAACAGGTGGCCAAATC	267
Qy	241	AGACAGGGAGCTAGCCACATCTCCAAAACACAAGTCACTGAGCTATTGTGAACAGACT	300
Db	268	AGACAGGGAGCTAGCCACATCTCCAAAACACAAGTCACTGAGCTATTGTGAACAGACT	327
Qy	301	CGACCGTCATGACAGCTGTGACAGATTTTAATAATCTCCGGGACAGATTTAAAGACTTC	360
Db	328	CGACCGTCATGACAGCTGTGACAGATTTTAATAATCTCCGGGACAGATTTAAAGACTTC	387
Qy	361	TCCTGATTGAGATGCCGGGAATTAATCTGCTGCTTATAGATCAACCGAAGAAATGTGAA	420
Db	388	TCCTGATTGAGATGCCGGGAATTAATCTGCTGCTTATAGATCAACCGAAGAAATGTGAA	447
Qy	421	AGAAGCCCTCTTACTATTGGCGATAAAGTCCCTCAITCCACATTTTCCCAAGAGGCAAG	480
Db	448	AGAAGCCCTCTTACTATTGGCGATAAAGTCCCTCAITCCACATTTTCCCAAGAGGCAAG	507
Qy	481	CAATTCCGGAGTCTTGGGGCGGAGAACCAAGTAGGAGAACACAGACAGTACTGAGGGTCT	540
Db	508	CAATTCCGGAGTCTTGGGGCGGAGAACCAAGTAGGAGAACACAGACAGTACTGAGGGTCT	567
Qy	541	TCCTGTTGGGCAAGACACCCCGAGAGAACACCCCTGACCTTTCGGACATGCTTAAGT	600
Db	568	TCCTGTTGGGCAAGACACCCCGAGAGAACACCCCTGACCTTTCGGACATGCTTAAGT	627
Qy	601	TTGAGAGTGACAAGCACCCAGGACATCTCATGTGGAACATATAGACACACATCTTCAACC	660
Db	628	TTGAGAGTGACAAGCACCCAGGACATCTCATGTGGAACATATAGACACACATCTTCAACC	687
Qy	661	TGTCCTCTGAAGGAAGTGTGTTCTTAGTGGTGAGCAGCTTCTGTCACACAGCAGAGT	720
Db	688	TGTCCTCTGAAGGAAGTGTGTTCTTAGTGGTGAGCAGCTTCTGTCACACAGCAGAGT	747
Qy	721	TTGCTCTCAAGGGCGATGATGACGTGTTTGTGAACACCCATCAGATCTTAATTAATTGGA	780
Db	748	TTGCTCTCAAGGGCGATGATGACGTGTTTGTGAACACCCATCAGATCTTAATTAATTGGA	807
Qy	781	ATAGCTTTATCCAAAGAGAAAGCCAAAGACTTGTTCATAGGTGAGTGTATCCACAATGCTG	840
Db	808	ATAGCTTTATCCAAAGAGAAAGCCAAAGACTTGTTCATAGGTGAGTGTATCCACAATGCTG	867
Qy	841	GGCCTCACCGGATGAAGAACTGAAGTACTACATCCCAAGAGTCTTCTACACCGCGTCT	900
Db	868	GGCCTCACCGGATGAAGAACTGAAGTACTACATCCCAAGAGTCTTCTACACCGCGTCT	927
Qy	901	ACCCACCGTATGCCGGGGGTGGTGGATTCTCTACTCCGGCCCCCTTGCCTTTAGGGTGT	960
Db	928	ACCCACCGTATGCCGGGGGTGGTGGATTCTCTACTCCGGCCCCCTTGCCTTTAGGGTGT	987
Qy	961	ACAGTGCAGCTAGCCGGGTCCATCTCTACCTATTGATGATGTTTATACGGGAATGTGCC	1020
Db	988	ACAGTGCAGCTAGCCGGGTCCATCTCTACCTATTGATGATGTTTATACGGGAATGTGCC	1047

QY 1021 TTCAGAACTGGCCCTTGTTCCAGAGAGACACAAAGCTTCAGGACATTTGATATTGAAG 1080
 |||||||
 Db 1048 TTCAGAACTGGCCCTTGTTCCAGAGAGACACAAAGCTTCAGGACATTTGATATTGAAG 1107
 |||||||
 QY 1081 AGAAATAAGAAATAATTTGCTCTATATAGACCTAAATGTTAGTACATACAGAAAC 1140
 |||||||
 Db 1108 AGAAATAAGAAATAATTTGCTCTATATAGACCTAAATGTTAGTACATACAGAAAC 1167
 |||||||
 QY 1141 CTCAGAGATGATGATATCTGCTCAGTTCAGTTCAGAAAGTCTCTAATTTAAATCTGA 1196
 |||||||
 Db 1168 CTCAGAGATGATGATATCTGCTCAGTTCAGTTCAGAAAGTCTCTAATTTAAATCTGGA 1223
 |||||||

RESULT 5
 AC119549/c
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-444D1, WORKING DRAFT SEQUENCE, 2
 unordered pieces.
 AC119549
 AC119549.5 GI:25008693
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE Rattus norvegicus
 ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 178009)
 Muzny, D., Maric, Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 George, G., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hayes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Iglebird, D., Jackson, A.,
 Jackson, L., Jacob, S., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensu, L., Louisedge, H., Lozado, R. J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindratne, M., Mahmoud, M., Mailloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mahoney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nait, L.,
 Narkar, C., Neal, D., Newton, N., Ngweni, N., Norris, S.,
 Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, I., Prannkoc, C.,
 Pioppe, F., Poindexter, A., Popovic, D., Pirmus, E., Pu, L.,
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,
 Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczka, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,

Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, S., Dunn, D., von
 Niederhauser, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
 Weinstock, G. and Gibbs, R. A.
 Direct Submission
 Unpublished
 2 (bases 1 to 178009)
 Worley, K. C.
 Direct Submission
 Submitted (28-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 178009)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 15, 2002 this sequence version replaced gi:23194601.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GWDO
 Center clone name: CH230-444D1
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 156182 bases at least Q40
 Consensus quality: 157889 bases at least Q30
 Consensus quality: 159078 bases at least Q20
 Estimated insert size: 161372; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 175629: contig of 175629 bp in length
 * 175630 175729: gap of unknown length
 * 175730 178009: contig of 2280 bp in length.
 FEATURES
 source
 1. 178009
 Location/Qualifiers
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-444D1"
 785..1598
 /note="clone_boundary
 clone_end:T7
 site:
 end_sequence=BZ198810"
 complement(166970..167869)
 /note="clone_boundary
 clone_end:Sp6
 misc_feature
 misc_feature
 complement(166970..167869)
 /note="clone_boundary
 clone_end:Sp6

```

Site:
end_sequence:BZ198831*
misc_feature 16959..172426
/note="wgs_end_extension
clone_end:Sp6*
misc_feature 17316..175629
/note="wgs_end_extension
clone_end:Sp6*

BASE COUNT 43995 a 38676 c 36593 g 40919 t 17826 others
ORIGIN

Query Match 89.0%; Score 1064.4; DB 2: Length 178009;
Best Local Similarity 93.2%; Pred. No. 7.8e-289;
Matches 1113; Conservative 0; Mismatches 81; Indels 0; Gaps C;

QY 1 AGATGAGTGTGGGGCGTCAAGAGTCAAGTTCCTGGGCAATCCTGATGATGCAAAATGCT 60
Db 144675 AAATGAGTGTGGGGCGTCAAGAGTCAAGTTCCTGGGCAATCCTGATGATGCAAAATGCT 144615

QY 61 TCATTTATTTGATTTGGAAGTCTCCAAAACAGTAGCCCAAGACAAAATGGGAAGGAG 120
Db 144615 TCATTTATTTGATTTGGAAGTCTCCAAAACAGTAGCCCAAGACAAAATGGGAAGGAG 144556

QY 121 GAGTAAATATCCGAAAGAGAGTTCCTGGAAGCCACCCAGCACCTCCCGGGGCATACATGGA 180
Db 144555 GAGTAAATATCCGAAAGAGAGTTCCTGGAAGCTTCCTGAGCTTCCTCCGCGCATACATGGA 144496

QY 181 ACAGGGAACAGAGAGTGTGACAGGTGGTACATCCCATCTTGAACAGGGGTGCCAATC 240
Db 144495 ATAGGAGCAGGAGAGTGTGACAGGTGGTACATCCCATCTTGAACAGGGGTGCCAATC 144436

QY 241 AGACAGGGGAGCTAGCCATCTCCAAACACAGTCACTGAGTATTGTGAACAGACACI 300
Db 144435 AGACAGGGGAGCTGTTCATCTCCAAACACAGTCACTGAGTATTGTGAACAGGACI 144376

QY 301 CGAGCGTCAATGACAGTGTGACAGATTTTAATAATCTCCGAGACAGATTTAAAGACTTC 360
Db 144375 CGAGCTCATGACAGTGTGACAGATTTTAATAATCTCCGAGACAGATTTAAAGACTTC 144316

QY 361 TCTTGTATTGAGATCGCGAATTAATCTCGCTGCTATAGATCAACCGAAGAAATGTGCA 420
Db 144315 TCTTGTATTGAGATCGCGAATTAATCTCGCTGCTATAGATCAACCGAAGAAATGTGCA 144256

QY 421 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAG 480
Db 144255 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTCCAGAGGCAAG 144195

QY 481 CAATTCGGAGTCTTGGGCGCCAGAACCAACGCTAGGAAACAGACATAGTGGGTCT 540
Db 144195 CAATTCGGGAATCTTGGGCGCCAGAGACCAATGTAGGAAACAGACATAGTGGGTCT 144136

QY 541 TCTGTGGGCAAGACACCCCGAGGAGCAACACCCCTGACCTTTCGGACATGCTTAAGT 600
Db 144135 TCTGTGGGCAAGACACCCCGAGGAGCAACACCCCTGACCTTTCGGACATGCTGAAT 144076

QY 601 TTGAGAGTCAAGACACACAGGACATCTCTATGTGGAATATAGACACATCTTCAACC 660
Db 144075 TCAGAGTGAAGAGCAGGACATCTCTATGTGGAATATAGACACATCTTCAACC 144015

QY 661 TGTCCCTGAAGAGTGTGTTTCTTAGTGGGTCAGGACATCTCTGTCACACACAGT 720
Db 144015 TGTCTCTCAAGGAAGTACTGTTCTCAGTGGGTGAGCACCTCTCTGTCACACACAGT 143955

QY 721 TTGCTTCAAGGCGATGATGAGTGTGTTGGAACACCCATCACATCTCTTAATCTGA 780
Db 143955 TTGCTTCAAGGCGATGATGAGTGTGTTGGAACACCCATCACATCTCTGAATCTGA 143896

QY 781 ATAGCTATTCAAGACCAAGCAAGACTTTTTCATAGGTGACGTGATCCCAATGCTG 840
Db 143895 ATAGCTATTCAAGACCAAGCAAGACTTTTTCATAGGTGACGTGATCCCAATGCTG 143836

QY 841 GGCCTCACCGGATAGAACTGAGTACTACATCCCAAGACTCTTCTACACCGCGTCT 900
Db 143836 GGCCTCACCGGATAGAACTGAGTACTACATCCCAAGACTCTTCTACACCGCGTCT 143776

Db 143835 GGCCTCACCGGATAGAACTGAGTACTACATCCCAAGACTCTTCTACACCGCGTCT 143776

QY 901 ACCACCGTATCCCGGGGTGGGGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Db 143775 ACCACCGTATCCCGGGGTGGGGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 143716

QY 961 ACAGTGCAGTACGCGGGTCCATCTCTACCTTATGATGATGATGATGATGATGATGATGAT 1020
Db 143715 ACAATGTAATGACCGGGTCCACCTCTACCTTATGATGATGATGATGATGATGATGAT 143656

QY 1021 TTCAGAACTGGGCTTGTTCAGAGAGACCAAGGCTTCAGGACATTTGATTTGAAG 1080
Db 143655 TTCAGAACTGGGCTTGTTCAGAGAGACCAAGGCTTCAGGACATTTGATTTGAAG 143596

QY 1081 AGAAAAATAGAAAAATATTTTCTCTATATAGACTTAATGTTAGTACATACAGAAAAAC 1140
Db 143595 AGAAAAATAGAAAAATATTTTCTCTATATAGACTTAATGTTAGTACATACAGAAAAAC 143536

QY 1141 CTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1194
Db 143535 CTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 143482

AC109547 318862 bp DNA linear HTG 11-OCT-2002
Rattus norvegicus clone CH230-217N7, *** SEQUENCING IN PROGRESS
***, 3 unordered pieces.
AC109547
AC109547.4 GI:23603983
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 318862)
Muzny,D,Marie,,Metzker,M,Lee,,Abramzon,S,,Adams,C,,Alder,J,,
Allen,C,,Allen,H,,Alsbrooks,S,,Amin,A,,Anguiano,D,,
Anyalebechi,V,,Ayodeji,M,,Baca,E,,Baden,H,,
Baldwin,D,,Bandaranaika,D,,Barber,M,,Barnstead,M,,Benahmed,F,,
Biswal,K,,Blair,J,,Blankenburg,K,,Blyth,P,,Brown,M,,
Bryant,N,,Buhay,C,,Burch,P,,Burrell,K,,Calderon,E,,
Cardenas,V,,Carter,K,,Cavazos,I,,Ceasar,H,,Center,A,,
Chacko,J,,Chavez,D,,Chen,G,,Chen,R,,Chen,Y,,Chen,Z,,Chu,J,,
Cleveland,C,,Cockrell,R,,Cox,C,,Coyle,M,,Cree,A,,D'Souza,L,,
Davila,M,L,,Davis,C,,Davy-Carroll,L,,De Anda,C,,Dederich,D,,
Delgado,O,,Denson,S,,Deramo,C,,Ding,Y,,Dinh,H,,Divya,K,,
Draper,H,,Dugan-Rocha,S,,Dunn,A,,Durbink,K,,Duval,B,,Eaves,K,,
Egan,A,,Escotto,M,,Eugene,C,,Evans,C,A,,Falls,T,,Fan,G,,
Fernandez,S,,Finley,M,,Flagg,N,,Forbes,L,,Foster,M,,Foster,P,,
Fraser,C,M,,Gabisi,A,,Ganta,R,,Garcia,A,,Garner,T,,Garza,M,,
Gebregeorgis,E,,Geer,K,,Gill,R,,Grady,M,,Guerra,W,,Guevara,W,,
Gunarathne,P,,Haaland,W,,Hamil,C,,Hamilton,C,,Hamilton,K,,
Harvey,X,,Havlak,P,,Hawes,A,,Henderson,N,,Hernandez,J,,
Hernandez,R,,Hines,S,,Hladun,S,L,,Hodgson,A,,Hogues,M,,
Hollins,B,,Howells,S,,Hulyk,S,,Hume,J,,Idlebird,D,,Jackson,A,,
Jackson,L,,Jacob,L,,Jiang,H,,Johnson,B,,Johnson,R,,Jollivet,A,,
Karpathy,S,,Kelly,S,,Khan,Z,,Khan,Z,,King,L,,Kovar,C,,
Kowals,C,,Kraft,C,L,,Lebow,H,,Levan,J,,Lewis,L,,Li,Z,,Liu,J,,
Liu,J,,Liu,W,,Liu,Y,,London,P,,Longacre,S,,Lopez,J,,
Lorenshewa,L,,Loulsegh,H,,Lozard,R,J,,Lu,X,,Ma,J,,
Maheshwari,M,,Mahindartne,M,,Mahmoud,M,,Malloy,K,,Mangum,A,,
Mangum,B,,Mapua,P,,Martin,K,,Martin,R,,Martinez,E,,
Mawhinney,S,,McLeod,M,P,,McNeill,T,Z,,Meenen,E,,
Milosavljevic,A,,Miner,G,,Minja,E,,Montemayor,J,,Moore,S,,
Morgan,M,,Morris,K,,Morris,S,,Munidas,M,,Murphy,M,,Nair,L,,
Nankervis,C,,Neal,D,,Newton,S,,Nguyen,N,,Norris,S,,
Nwaokelimeh,O,,Okwuonu,G,,Olanpunsagoo,A,,Pal,S,,Parks,K,,
Pasternak,S,,Paul,H,,Perez,A,,Perez,L,,Pfankoch,C,,
Plopper,F,,Polinder,J,,Popovic,D,,Primus,E,,Pu,L,,
Puzos,M,,Quiroz,J,,Rachlin,E,,Reeves,K,,Regier,M,A,,Reigh,R,,
Reilly,B,,Reilly,M,,Ren,Y,,Reuter,M,,Richards,S,,Riggs,F,,

```

Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Sma's, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, I., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, H., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhauser, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 318862)
Worley, K.C.

Direct Submission
Submitted (05-PEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 318862)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Oct 9, 2002 this sequence version replaced gi:21738004.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GQCH
Center clone name: CH230-217N7
----- Summary Statistics
Assembly program: Phrap; version 0.90329
Consensus quality: 188785 bases at least Q40
Consensus quality: 192484 bases at least Q30
Consensus quality: 195081 bases at least Q20
Estimated insert size: 195808; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.htm).
* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
* 1 314235: contig of 314235 bp in length
* 314236 314335: gap of unknown length
* 314336 315730: contig of 1395 bp in length
* 315731 315830: gap of unknown length
* 315831 318862: contig of 3032 bp in length.
Location/Qualifiers
1. .318862

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-217N7"
30004..31805
/note="vgs_end_extension
clone_end:T7"
84302..85148
/note="clone_boundary
clone_end:T7
site:ECORI
end_sequence:RWBMC76TJ"
262652..263167
/note="clone_boundary
clone_end:sp6
site:ECORI
end_sequence:RWBMC76TV"
BASE COUNT 52536 a 45389 c 45802 g 54010 t 121125 others
ORIGIN

Query Match 88.3%; Score 1056.4; DB 2; Length 318862;
Best Local Similarity 92.8%; Pred. No. 1.6e-286;
Matches 1108; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 1 AGATGAGTGTGGGGCGTCGAAGAGTCAAGTGTCTGGGCATCCTGTGATGGCAATGTCT 60
DB 300006 AAATGAGTGTGGGGCGTCGAAGAGTCAAGTGTCTGGGCATCCTGTGATGGCAATGTCT 300065

QY 61 TCATTTATTTGATTTGTGGAAGTCTCCAAAACAGTAGCCAAAGACAAAATGGAAGGGAG 120
DB 300066 TCATCTATTTGATTTGTGGAAGTCTCCAAAACAGTAGCCAAAGACAAAATGGAAGGGAG 300125

QY 121 GAGTAATAATCCCAAGAAAGTTCCTGGAAGCCACCACACTCCCGGGCATACTGGA 180
DB 300126 GAGTGATATAACCAAGAGAGTTCCTGGAAGTTCCTCCCGGGCATACTGGA 300185

QY 181 ACAGGGAACAGAGAGAGTGAACAGTGTGTACAAATCCCATCTTTGACAGGGTGGCCAATC 240
DB 300186 ATAGGGAGCAGAGAGAGTGAACAGTGTGTACAAATCCCATCTTTGACAGGGTGGCCAATC 300245

QY 241 AGACAGGGGAGCTAGCCACATCTCCAAACACAGTCACTGAGCTATTGTGAACACAGACT 300
DB 300246 AGACGGGGGATCTGTTCAATCTCCAAACACAGTCACTGAGCTATTGTGAACACGACT 300305

QY 301 CGACGGTCATGACAGCTGTGACAGATTTTAAATATCTGCCGACAGATTTTAAAGACTTTC 360
DB 300306 CGACAGTCATGACAGCTGTGACAGATTTTAAATATCTGCCGACAGATTTTAAAGACTTTC 300365

QY 361 TCTTGTATTGAGATGCGGGAATTTACTCGCTTATAGATCAACCGAAGAAATGTGCAA 420
DB 300366 TCTTGTATTGAGATGCGGGAATTTACTCGCTTATAGATCAACCGAAGAAATGTGCAA 300425

QY 421 AGAAGCCCTTCTTACTATITGGCGATAAAGTCCCTCATTTCCACATTTTGGCAGAGCAAG 480
DB 300426 AGAAGCCCTTCTTACTATITGGCGATAAAGTCCCTCATTTCCACATTTTGGCAGAGCAAG 300485

QY 481 CAATTCGGGAGTCTTGGGGCGGAGAACACCAACAGTAGGGAACACAGAGTAGTGAGGCTCT 540
DB 300486 CAATTCGGGAGTCTTGGGGCGGAGAACCAACAGTAGGGAACACAGAGTAGTGAGGCTCT 300545

QY 541 TCCTGTTGGGCAAGACACCCCGAGGACCAACCACTGACCTTTTCGGACATGCTTTAAGT 600
DB 300546 TCCTGTTGGGCAAGACACCCCGAGGACCAACCACTGACCTTTTCGGACATGCTTTAAGT 300605

QY 601 TTGAGAGTGACAGCAGACATCTCTCATGTGGAACTATAGAGACACATTTCTCAACC 660
DB 300606 TCGAGAGTGAGAGCAGCAGACATCTCTCATGTGGAACTATAGAGACACATTTCTCAACC 300665

QY 661 TGTCCTCAAGGAAGTCTGTTTCTTAGTGGGTGACACTTCTGTCAGAGACGAGACT 720
DB 300666 TGTCCTCAAGGAAGTCTGTTTCTTAGTGGGTGACACTTCTGTCAGAGACGAGACT 300725

QY 721 TTGCTTTCAAGGGGATGATGACGCTGTTTGTGAACACCCATCATCTCTTAATTACTTGA 780

FEATURES
source

Db	300726	TTGCTTCAAGGGTGATGATGAGCTGTCTGTGAAACCCATCACATCCTGAATACTTGA	300785
Qy	781	ATAGCTTATCCAGAGCAAGCAAGAACTGTTTCATAGTGACGTGATCCACAAATGCTG	840
Db	300786	ATAGCTTATCCAGAGCAAGCAAGAACTGTTTCATAGTGACGTGATCCACAAATGCTG	300845
Qy	841	GGCTCACCGGATAGCAAACTGAAGTACTACATCCAGAACTCTCTACACCGGGTCT	900
Db	300846	GGCCCCATCGGGATAGCAAACTGAAGTACTACATCCAGAACTCTCTACACCGGGTCT	300905
Qy	901	ACCACCGTATGCCGGGGTGGTGGATTCTCTGACTCCGGCCCCCTTGCCTTGAAGGCTG	960
Db	300906	ACCACCGTATGCCAGGGGGTGGTGGTCTCTGACTCCGGCCCCCTTGCCTTGAAGGCTT	300965
Qy	961	ACAGTGGCATAGCCGGGTGCATCTCTACCCATTTGATGATGTTTATACGGGAATGGCC	1020
Db	300966	ACATGTAAGTGAACCGGGTCCACCTCTACCCATTCGATTAACGCTCAATCTGGAATGGCC	301025
Qy	1021	TTCAGAAACTGGGCTTGTTCAGAGCAAGCAAGGCTTCAGACATTTGATATTGAAG	1080
Db	301026	TTCAGAAACTGGGCTTGTTCAGAGCAAGCAAGGCTTCAGACATTTGATATTGAAG	301085
Qy	1081	AGAAAAATAGAAAAATATTGTTCTCATATAGACCTAAAGTTAGTACATAGCAGAAAAAC	1140
Db	301086	AGAAAAATAGAAAAATATTGTTCTCATATAGACCTAAAGTTAGTACATAGCAGAAAAAC	301145
Qy	1141	CTCAGAGATGATGATATCTGCTCAGTTCAGTGCAGATCTTAATTAAGTGT 1194	
Db	301146	CTCAGAGATGATGATATCTGCTCAGTTCAGTGCAGATCTTAATTAAGTGT 301199	
RESULT 7			
AF092051	1831 bp mRNA linear: PRI 09-AUG-2000		
LOCUS	Homo sapiens beta-1,3-N-acetylglucosaminyltransferase mRNA,		
DEFINITION	complete cds.		
ACCESSION	AF092051		
VERSION	AF092051.2 GI:9755416		
KEYWORDS	.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 1831)		
	Zhou,D., Dinter,A., Gutierrez Gallego,R., Kamerling,J.P.,		
JOURNAL	Viegant, J.F., Berger, E.G. and Hennet, T.		
	A beta-1,3-N-acetylglucosaminyltransferase with		
MEDLINE	poly-N-acetylglucosamine synthase activity is structurally related		
	to beta-1,3-galactosyltransferases		
PUBMED	Proc. Natl. Acad. Sci. U.S.A. 96 (2), 406-411 (1999)		
REFERENCE	99110903		
AUTHORS	2 (bases 1 to 1831)		
TITLE	Zhou,D., Berger,E.G. and Hennet,T.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (15-SEP-1998) Physiology, University of Zurich,		
	Winterthurerstrasse 190, Zurich 8057, Switzerland		
AUTHORS	3 (bases 1 to 1831)		
TITLE	Zhou,D., Berger,E.G. and Hennet,T.		
JOURNAL	Direct Submission		
REMARK	Submitted (09-AUG-2000) Physiology, University of Zurich,		
	Winterthurerstrasse 190, Zurich 8057, Switzerland		
COMMENT	Sequence update by submitter		
FEATURES	On Aug 9, 2000 this sequence version replaced gi:4191393.		
source	Location/Qualifiers		
	1..1831		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/tissue_type="brain"		
	/dev_stage="fetus"		
	236..1429		
CDS			

/EC_number="2.4.1.149"			
/note="glycosyltransferase; b3Gnt"			
/codon_start=1			
/product="beta-1,3-N-acetylglucosaminyltransferase"			
/protein_id="AAD09764.2"			
/db_xref="GI:9755417"			
/translation="MSVGRRRRIKLILMMANVFYFIMEVSKSSQSKNGKGEVIIP			
KEKFKISTPPEAYWNRQEKLRNYNPILSMLTNOTGAGRLSNISHLNCEPLRV			
TSVVTGFNNLPDRKDFLLYLCRNYSLIDQDKCAKPFLLLAIKSLTFHARRQA			
IRESGQESNAGNQTWRVFLGQTPPDNHPDLGMLKFESEKHODILMNNYRDTFF			
NLSLVELFLRWVSTSCDPTFEVFGDDDFVFNTHILNLSLKTAKDLFDGDI			
HNAGSRHDKLKLKYYIPEVYVSGLYPEVAGGGGLYSGHLALRLYHTDOVHLYPDDV			
YTGMLCKLGLVPEKHKGFTEDIEEKNKNNICSVYDMLVHSPRQEMIDINWSQLQS			
AHLKC"			
BASE COUNT	497 a	423 c	443 g 468 t
ORIGIN			
Query Match	75.8%; Score 906.4; DB 9; Length 1831;		
Best Local Similarity	84.9%; Pred. No. 1.5e-244;		
Matches :015; Conservative	0; Mismatches 181; Indels 0; Gaps 0;		
Qy	1	AGATGAGTGTGGGGGTGCAAGAGTCAAGTGTCTGGGCATCCTGATGATGGCAAAATGCT	60
Db	234	AAATGAGTGTGGACGTCCAGAAATAAAGTGTGTGGGTATCTGATGATGGCAAAATGCT	293
Qy	61	TCATTATTGATGTGGAGTCTCCAAAACAGTAGCCAAACAAAATGAAAGGGAG	120
Db	294	TCATTATTATTATGGAAGTCTCCAAAAGCAGTAGCCAAAGAAAATGAAAGGGG	353
Qy	121	GAGTAATAATCCGAAAGAGAGTCTGGAAGCCACCCAGCACCTCCCGGGCATACTGA	180
Db	354	AGTAATAATATCCCAAGAGAGTCTGGAAGATATCTACCCCTCCCGAGGCACTGGA	413
Qy	181	ACAGGAAACAGAGAGTGAACAGTGGTACAAATCCCATCTTGAACAGGGTGGCAATC	240
Db	414	ACCGAGCAAGAGAGAGTGAACCGGCAGTACAAACCCCATCTGAGCATGCTGACCAAC	473
Qy	241	AGACAGGGGAGTGCACACATCTCCAAACAGTAGTCACTGAGCTATGTGAACAGACT	300
Db	474	AGACGGGGAGGGCGGCGAGGCTCTCCAAATATTAAGCCATCTGAACCTACTCGAACCT	533
Qy	301	CGACGGTCAATGACAGCTGTGACAGATTTTAAATAATCTGCCGGACAGATTAAGACTTC	360
Db	534	TGAGGGTCACTGGTGTGAGGGTTTAAACAATCTGCCGGACAGATTAAGACTTC	593
Qy	361	TCCTGTATTGAGATGCCGGAATTTACTCGCTGCTTATAGATCAACCGAAGAATGTCAA	420
Db	594	TGCTGTATTGAGATGCCGCAATTTATTCACCTGCTTATAGATCAGCGGATAAGTGTCAA	653
Qy	421	AGNAGCCCTCTTACTATTGGGGATAAAGTCCCTCATTCACATTTTGGCAGAGGCAAG	480
Db	654	AGAAACCTTTCTGTGTGGGATTAAGTCCCTCACTCCACATTTTGGCAGAGGCAAG	713
Qy	481	CAATCGGGAGTCTTGGGGCCGAGAAACCAACGTAGGGAACCAAGAGTGTGGGTCT	540
Db	714	CAATCGGGAATCTTGGGGCCGAGAAACCAACGTAGGGAACCAAGAGTGTGGGTCT	773
Qy	541	TCTGTGTGGGCAAGACACCCCGAGAGCAACCCCTGACCTTTTGGACATGCTTAAGT	600
Db	774	TCTGTGTGGGCAAGACACCCCGAGAGCAACCCCGACCTTCAGATATGCTGAAAT	833
Qy	601	TTGAGAGTACAAAGCACACAGGACATCCTCATGTGAACTATAGAGACATCTTCAACC	660
Db	834	TTGAGAGTACAAAGCACACAGGACATCTTATGTGGAATACAGACACATTTCTCAAT	893
Qy	661	TGTCCTGAAGAGTGTGTTTCTTAGTGGTGGAGCACTTCTGTCAGAGGAGT	720
Db	894	TGTCCTGAAGAGTGTGTTTCTTAGTGGTGGAGCACTTCTGTCAGAGGAGT	953
Qy	721	TTGCTTCAAGGGCGATGATGACGTGTTTGTGAACACCCATCACATCTCTTAATTTACTGA	780
Db	954	TTGCTTCAAGGGCGATGACGTGTTTGTGAACACCCATCACATCTCTGAATTTACTGA	1013

QY	781	ATAGCTTATCCAGAGCAAGCCAAAGACTTGTTTCATAGTGACGTGATCCACAAAGCTG	843
Db	1014	ATAGTTTATCAAGACCAAGCCAAAGACTCTCTTCATAGTGATGATCCACAAATGCTG	1073
QY	841	GGCCTCACCGGTAAGCAAACTGAAGTACTACATCCCAAGTCITCTACACGGCGTCT	900
Db	1074	GACCTATCGGTAAGCAAACTGAAGTACTACATCCCAAGTCITCTTCCTGCGCTCT	1133
QY	901	ACCCACCGTATGCGGGGGTGGTGGATTCCTGTACTCCGCGCCCTTGCTTGAGCGTCT	960
Db	1134	ACCCACCGTATGAGGGGGAGGGGGTTCCTTACTCCGCGCCACCTGGCGCTGAGCTGT	1193
QY	961	ACAGTGGACTAGCCGGTCCATCTCTACCCCTATTCATGATGATGTTTATACGGGAATG	1020
Db	1194	ACCATATCACTGACCGAGTCCATCTCTACCCCTATTCATGATGATGTTTATACGGGAATG	1253
QY	1021	TTCAAGAACTGGGCTTGTTCAGAGAAAGCACAAGGCTTCAGGACATTTGATATTGAAG	1080
Db	1254	TTCAAGAACTGGGCTTGTTCAGAGAAAGCACAAGGCTTCAGGACATTTGATATTGAAG	1313
QY	1081	AGAAAATAGAAAATATTTTGTCTATATAGACCTAATGTTAGTACATAGCAGAAAC	1140
Db	1314	AGAAAATAGAAAATATTTTGTCTATATAGACCTAATGTTAGTACATAGTACAGAAAC	1373
QY	1141	CTCAAGAGATGATTGATATCTGGTCTCAGTTGCAAGTCCATAATTTAAAGTCTGA	1196
Db	1374	CTCAAGAGATGATTGATATTGGTCTCAGTTGCAAGTCCATAATTTAAAGTCTGA	1429
RESULT 8			
BD093191		1912 bp	DNA linear PAT 27-AUG-2002
LOCUS	BD093191		
DEFINITION	Useful polypeptide.		
ACCESSION	BD093191		
VERSION	BD093191.1 GI:22638779		
KEYWORDS	WO 0100848-A/1.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (bases 1 to 1912)		
	Sasaki,K., Shiraishi,N., Natsume,A., Yamada,Y., Nakagawa,S. and		
	Sekine,S.		
TITLE	Useful polypeptide		
JOURNAL	Patent: WO 0100848-A 1 04-JAN-2001;		
	KYOWA HAKKO KOGYO CO LTD,KATSUTOSHI SASAKI,NORIHIKO SHIRAISHI,		
	AYUMI NATSUME,YOJI YAMADA,SATOSHI NAKAGAWA,SUSUMU SEKINE		
COMMENT	OS Homo sapiens (human)		
	PN WO 0100848-A/1		
	PD 04-JAN-2001		
	PE 29-JUN-2000 WO 2000JP004304		
	PR 29-JUN-1999 JP 99p 183437,16-MAR-2000 JP OCP 074757 PI		
	KATSUTOSHI SASAKI,NORIHIKO SHIRAISHI,AYUMI		
	NATSUME,YOJI YAMADA.		
	PI SATOSHI NAKAGAWA,SUSUMU SEKINE		
	PC C12N15/54,C12N9/10,C12N5/10,C12N1/21,A61K48/00,A61K45/00, PC		
	A61K39/395,		
	PC A61K35/00,A61K31/711,A61P35/00,A61P29/00,A01K67/027,A01H5/00,		
	PC G01N33/53,		
	PC G01N33/15,G01N33/05		
	CC		
FEATURES	Key Location/Qualifiers.		
source	Location/Qualifiers		
	1..1912		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
BASE COUNT	518 a 430 c 457 g 507 t		
ORIGIN			
Query Match	75.8%; Score 906.4; DB 6; Length 1912;		
Best Local Similarity	84.9%; Pred. No. 1.5e-244;		
Matches 1015; Conservative	0; Mismatches 181; Indels 0; Gaps 0;		

QY	1	AGATGAGTGTGGCGCTGCAAGAGTCAAGTTGCTGGCATCTCTGATGATGCGAAATGTCT	60
Db	232	AAATGAGTGTGGACGTGCAAGATAAAGTTGTGGGTATCCTGTGATGCGAAATGTCT	291
QY	61	TCATTTATTGATTTGTGGAAGTCTCCAAAGACAGTAGCCAAAGACAAAATGGAAGGGGAG	120
Db	292	TCATTTATTGATTTATGGAAGTCTCCAAAGACAGTAGCCAAAGACAAAATGGAAGGGG	351
QY	121	GAGTAATAATCCCAAGAGAGAAAGTTCTGGAAGCCACCCAGCACTCCCGGGCATACTGGA	180
Db	352	AAGTAATAATCAACCAAGAGAGAGTTCTGGAAGATATCTACCCCTCCCGGGCATACTGGA	411
QY	381	ACAGGGAACAGAGAGAGCTGAACAGTGGTACAAATCCATCTTTGAACAGGCTGGCCAAATC	240
Db	412	ACCAGAGCAAGAGAGAGCTGAACCGGCAGTACAAACCCATCTCTGAGCATCTGACCAACC	471
QY	241	AGACAGGGGAGCTAGCCACATCTCCAAACACAAAGTCAAGTCAAGTATTTGGAACACAGACT	300
Db	472	AGACGGGGAGCGGGCAGGCTCTCCAAATTAAGCCATCTGAACACTCTCGAACCTTGACC	531
QY	301	CGACGGTCATGACAGCTGTGACAGATTTTAAATATCTGCCGGACAGATTTTAAAGACTTTC	360
Db	532	TGAGGTCACGTCGGTGTACGGGTTTAAACAACTTGGCGGACAGATTTTAAAGACTTTC	591
QY	361	TCCTGTATTTCAGATGCGGGAATTAAGTCTGCTTATAGATCAACGCAAGAAATGTGCA	420
Db	592	TGCTGTATTTCAGATGCGGCAATTAATTCACCTGCTTATAGATCAACGCGGATGATGTGCA	651
QY	421	AGAAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTTCCACATTTTGCAGAGCAAG	480
Db	652	AGAAAGCCCTTCTTCTGTCGCGATTAAAGTCCCTCACTCCACATTTTGCAGAGCAAG	711
QY	481	CAATTCGGGAGTCTTGGGGCGGAGAAACCAACAGTAGGGAACCAAGTAGTACAGGCTCT	540
Db	712	CAATTCGGGGAATCTGGGGCCCAAGAAAGCAACGAGGGAACCAAGCGTGTGCGAGTCT	771
QY	541	TCCTGTTGGGCAAGACACACCCCGAGGACCAACCCCTGACCTTTGCGACATGCTTAAAGT	600
Db	772	TCCTGCTGGCGGACACACACCCCGAGGACCAACCCCGAGCTTTCAGATATGCTGAAT	831
QY	601	TTGAGAGTGAACAGCACACAGGACATCTCTATGTGGAACACTATAGAGACACATTTCTCAACC	660
Db	832	TTGAGAGTGAAGACACCAAGACATTTCTATGTGGAACACTACAGAGACACTTTCTCAACT	891
QY	661	TGTCCTGAAGAAAGTCTGTTCTTAGTGGTGGAGCACTTCTGTCAGAGACGAGAGT	720
Db	892	TGTCCTGAAGAAAGTCTGTTCTCAGTGGGTAACTACTTCTGCGCCAGACACTGAGT	951
QY	721	TTGCTTCAAGGGGATGATGACGTGTTTGTGAACACCCATCATCTTAAATACTTGA	780
Db	952	TTGTTTCAAGGGGATGACGATGTTTGTGAACACCCATCATCTTAAATACTTGA	1011
QY	781	ATAGCTTATCAAGAGCAAGCCAAAGACTTGTTTCATAGTGAGCTGATCCCAATGCTG	840
Db	1012	ATAGTTTATCAAGAGCAAGCCAAAGACTCTCTTCAATAGTGATGATCCCAATGCTG	1071
QY	841	GGCCTCACCGGTAAGCAAACTGAAGTACTACATCCCAAGAGTCTTCTACACGGGCTCT	900
Db	1072	GACCTCATCGGTAAGCAAACTGAAGTACTACATCCCAAGAGTCTTCTACTCTGCGCTCT	1131
QY	901	ACCCACCGTATGCGGGGGTGGTGGATTCCTGTACTCCGCGCCCTTGCTTGAAGGCTGT	960
Db	1132	ACCCACCGTATGAGGGGGAGGGGGTTCCTTACTCCGCGCCACCTGGCGCTGAGGCTGT	1191
QY	961	ACAGTGGCACTAGCCGGTCCATCTCTACCTATTTGATGATGTTTATACGGGAATGTC	1020
Db	1192	ACCATATCACTGACCGAGTCCATCTCTACCCCTATGATGAGCTTTATCTGGAATGTC	1251
QY	1021	TTCAAGAACTGGGCTTGTTCAGAGAGAGCAAGGCTTCAGGACATTTGATATTGAAG	1080
Db	1252	TTCAAGAACTGGGCTTGTTCAGAGAGAGCAAGGCTTCAGGACATTTGATATTGAAG	1311


```

RESULT 10
BC030579
LOCUS
DEFINITION
Homo sapiens, UDP-GlcNAc:betaGal
beta-1,3-N-acetylglucosaminyltransferase 1, clone MGC:26071
IMAGE:4828158, mRNA, complete cds.
ACCESSION
BC030579
VERSION
1
KEYWORDS
GI:21040508
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2454)
AUTHORS
Strausberg, R.
TITLE
Direct Submission
JOURNAL
Submitted (13-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT
Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
Contact: amadan@systemsbio.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 34 Row: d Column: 20
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 15451893.
Location/Qualifiers
1. .2454
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="LocusID:10678"
/db_xref="taxon:9606"
/cclone="MGC:26071 IMAGE:4828158"
/tissue_type="Testis"
/cclone_lib="NIH_MGC_97"
/lab_host="DH108"
/notes="Vector: pBluescript"
113..1306
/codon_start=1
/product="UDP-GlcNAc:betaGal
beta-1,3-N-acetylglucosaminyltransferase 1"
/protein_id="AAH30579.1"
/db_xref="GI:21040508"
/translation="MSVGRRIKLLGLIMMANVFIYFMEVSKSSSEKNGKEVITP
KEKFWKTPPEAYNROEKLNRQYNPILSMLNTOTGEAGRLSNISHLNCEPLRV
TSVFTGNLDRKDFLLYLRGRNYSLLIDQPKCAKKPELLLAISLTPHARRQA
IRESQGSNAGNQTWVRVLLGOTPPEDNHPDLSMDLKFSEKHODILMWNVYRDTF
NLSLKEVLFKRWYSVCPDTEPVFKGDDDDVFNTHILNLSLTKAKDLFIQDVI
HNAGPHRDKLKIYIVVYSGLYPPYAGGGGFLYSGLHALRLRHITDQVHLYPIDV
YTGMLQKLGIVPEKHGKGFRTDIEKNKNICSYVDLMLVRSRKPQEMIDWSQLOS
AHLK"
BASE COUNT 725 a 444 c 524 g 761 t
ORIGIN
Query Match 75.8%; Score 906.4; DB 9; Length 2454;
Best Local Similarity 84.9%; Pred. No. 1.6e-244;
Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;
QY 1 AGATGAGTGTGGCGCTCGAAGAGTCAAGTTCGICGGGATCCTGAIGATGCGCAATGCTC 60
111 AAATGAGTGTGGCGCTCGAAGATAAAGTTGTTGGTATCTGATGATGCGCAATGCTC 170
61 TCATTTATTGATTTGGGAAGTCTCCAAACAGTAGCAAGCAAAATATGGAAGGAG 120
171 TCATTTATTGATTTGGGAAGTCTCCAAAGAGTAGCAAGCAAAATATGGAAGGAG 230
121 GAGTAATAATCCGGAAGAGAGTCTCGAAGCCACCCAGCACTCCCGGGCATCTGGA 180
231 AAGTAATAATACCCAAAGAGAGTCTCGAAGATATCTACCCCTCCCGAGGATCTGGA 290
181 ACAGGAACAGGAGAGAGTCTGAACAGGTGTAGCAATCCCATCTTGAACAGGGTGGCAATC 240
291 ACCGAGAGCAAGAGAGAGTCTGAACGCGACTACAAACCCATCTGAGCATGCTGACCAACC 350
241 AGACAGGGGAGTGTAGCCACATCTCCAAACAGTAGTCACTGAGCTATTGTGAAACAGACT 300
351 AGACGGGGAGGGCGGCGAGGCTCTCCAATATATAGCCATCTGAACACTGCGAACCTGACC 410
301 CGACGGTCTAGCAGCTGTGACAGATTTTAAATAATCTGCCGACAGATTTAAGACTTTC 360
411 TGAGGTCTAGCTCGGTGTTACGGGTTTAAACACTTGCAGCAGATTTAAGACTTTC 470
361 TCTTGTATTGAGATCCCGGAATTACTCGCTGCTTATAGATCAACCGAAGAAATGTGCAA 420
471 TGCTGTATTGAGATCCCGCAATTATTACTGCTTATAGATCAGCGGAGATAAGTGTGCAA 530
421 AGAAGCCCTCTTACTATTGGGAGTAAAGTCCCTCATTCACATTTTCCCAAGGCAAG 480
531 AGAACTCTTCTTGTGCTGGGAGTAAAGTCCCTCACTCCACATTTTGGCAGAGGCAAG 590
481 CAATCCGGAGTCTTGGGGCCGAGAAACCAAGCTAGGGAACAGACAGTAGTAGGGTCT 540
591 CAATCCGGGAATCTTGGGGCCAGAAAGCAACGCGAGGAAACCAAGGTTGTCGAGTCT 650
541 TCCTGTGGCAGACACACCCCGAGAGAGCAACCACTTACCTTTCGGACATGCTTAAGT 600
651 TCCTGTGGGCCAGACACCCCGAGAGGCAACCACTTTCAGATATGCTGAAAT 710
601 TTGAGAGTGTCAACAGCACCAGGACATCTCTGTGGAACTATAGAGACACATCTTCAACC 660
711 TTGAGAGTGAGAGACCAAGACATTTCTATGTGGAATACAGACACATCTTCTCAACT 770
661 TGTCCTCGAAGGAGTGTGTTTCTTAGTGGGTGAGCACTTCTCTGTCAGACAGCAGAGT 720
771 TGTCCTCGAAGGAGTGTGTTTCTCAGTGGGTAAGTACTTCTCTGCGCCAGACACTGAGT 830
721 TGTCCTCGAAGGCGGATGATGAGTGTGTTGTGAACACCCATCACATCCCTTAATTCGA 780
831 TGCTTTCAAGGCGGATGAGGATGTTTGTGNAACACCCATCACATCTCTGAAATTCGA 890
781 ATAGCTTATCCAAAGCAAGCAAGAGCTTGTTCATAGGTGACGTGATCCACAAATGCTG 840
891 ATAGTTATCCAAAGCAAGCAAGAGTCTCTTCATAGGTGATGTGATCCCAATGCTG 950
841 GGCTCACCCGGATAGAAACTGAACTACTACATCCCAAGAGTCTTCTACACCGCGCTCT 900
951 GACCTCATCGGATAGAAAGTGAAGTACTACATCCCAAGAGTGTGTTACTCTGCGCTCT 1010
901 ACCACCGATATCGCGGGTGGTGGATTCCTGTACTCCCGCCCCCTTGCTTGGAGGCTGT 960
1011 ACCACCGTATGCGGGGAGGGGGTCTCTTACTCCCGCCACCTGGCCCTGGAGGCTGT 1070
961 ACAGTCCGACTAGCCGGGTCCATCTCTACCCCTATTGATGATGTTTATAGCGGAATGTGCC 1020
1071 ACCATATCACTGACAGGTCCTCTACCCCTATGATGACGTTTATATCGAATGTGCC 1130
1021 TCCAGAACTGGCCCTTGTTCAGAGAGCAAGAGCTTCAGGACATTTGATATTTGAAG 1080
1131 TCCAGAACTGGCCCTGCTTCCAGAGAAACCAAGAGCTTCAGGACATTTGATATCGAGG 1190
1081 AGAAATAGAAATATTTGTTCCCTATATAGACCTAATGTAGTACATAGCAGAAAC 1140
1141

```


Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Hominidae: Homo.
1 (bases 1 to 2811)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schemper,C.F., Blatt,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raja,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muxny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S., Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalls,D.E., Schnerch,A., Schein,J.E., Jones,S.J., and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22389257
12477932
2 (bases 1 to 2811)
Strausberg,R.
Direct Submission
Submitted (03-MAR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgaps@email.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre.
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjeil, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 89 Row: 3 Column: 21
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 15451893.

FEATURES

source

1. .2811
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:49838 IMAGE:5767816"
/tissue_type="Brain, fetal, whole pooled"
/clone_lib="NIH_MGC_121"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1. .2811
/gene="B3GNT1"
/note="synonyms: BETA3GNT, B3GNT-2, B3GN-T2, B3GN-T1, B3GN1"
/db_xref="LocusID:10678"
/db_xref="MIM:605581"
239. .1432

gene

CDS

/codon_start=1
/product="beta-1,3-N-acetylglucosaminyltransferase bGNT-1"
/protein_id="AAH47933.1"
/db_xref="GI:28839675"
/db_xref="LocusID:10678"
/translation="MSVGRRIKLLGLIMANVFIYFIMEVSKSSQEKNGKEVILP
KEXFKISTPTPEAYWNRQELNKNQYNLSLNTOTGAGLSNLSHLYNCEPDLRV
TSVVTGNLPLDRFKDFLLVLCRNYSLLIDOPDKCAKPKFLLAYKSLTPHARRQA
IRSWGOESNAGNQTVVRVLLQOTPPDNHPLDSLAKFESEKHODILMWNRYRTFF
NLSLKEVFLRWVSTPCDFEYFVKGGDDVFNHILNLFNLSLSTKQDHLPIGDVI
HNSGPHDKLKYIIPVYVYSGGLYPPYAGGGFLYSGHLALRLYHTDQVHLPIGDV
YTMCLQKGLVPEKHKGFRTDIEEKNKNICSYVDLMVLVHRSKPOEMIDIWSOLQS
AHLKC"
BASE COUNT 837 a 527 c 613 g 834 t
ORIGIN
Query Match 75.8%; Score 906.4; DB 9; Length 2811;
Best Local Similarity 84.9%; Pred. No. 1.6e-244;
Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;
QY 1 AGATGAGTGTGGGGCGTCGAGAGTCAAGTTGCTGGGCATCCTGTGATGATGCGCAATGTCT 60
DB 237 AATGAGTGTGGGACGTCGAAGAATAAAGTTGTTGGGTATCTCTGATGATGCGCAATGTCT 296
QY 61 TCATTTATTTGATTGTGGAAGTCTCCAAAACAGTAGCCAGACAAAATAATGGAAGGGAG 120
DB 297 TCATTTATTTATTTATGGAAGTCTCCAAAAGCAGTAGCCAAAGAAAATAATGGAAGGGG 356
QY 121 GAGTAATAATCCGAAAGAGAAGTTCTTGGAGCCACCAGCACTCCCGGGGCATATCGGA 180
DB 357 AAGTAATAATACCCAAAGAGAAGTTCTTGGAGATATATCTACCCCTCCCGAGGACATCTGGA 416
QY 181 ACAGGGAACAGGAGAAGCTGAACAGGTGTACATCCATCCATCTTGAACAGAGGTGGCAATC 240
DB 417 ACCGAGCAACAGAGAAGCTGAACCGGCGAGTACAAACCCATCTCTGAGCATGCTGACCAACC 476
QY 241 AGACAGGGGAGCTAGCCACATCTCCAAACACAAAGTCACTGAGCTATTTGGAACCAAGACT 300
DB 477 AGACGGGGAGGGCGGCGGCTCTCCCAATATATAGCCATCTGAACTACTGCAACCTGACC 536
QY 301 CGACGGTCATGACAGCTGTGACAGATTTTAAATATCTGCCGGACAGATTTTAAAGACTTTC 360
DB 537 TGAGGGTCACGTCGGTGTACGGGTTTAAACAACTTGGCGGACAGATTTTAAAGACTTTC 596
QY 361 TCTTGTATTTCAGATGCGGGAATTTACTCGCTGCTTATAGATCAACCCGAAAGAAATGTGCAA 420
DB 597 TGCTGTATTTCAGATGCGGCAATTTATCACTGCTTATAGATCAGCGGAGATAGTGTCAA 656
QY 421 AGAAGCCCTCTTACTATTGCGGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAG 480
DB 657 AGAAACCTTTCTGTGCTGGCGATTAAAGTCCCTCATTCACATTTTCCAGAGGCAAG 716
QY 481 CAATTCGGGAGTCTTGGGGCGGAGAACCAACAGTAGGGAACACAGACAGTAGTGGAGTCT 540
DB 717 CAATCCGGGAATCTTGGGGCCAGAGACGAGCGAGGAGACCAACCGTGTGCGAGTCT 776
QY 541 TCCTGTTGGGCAACACACCCCGAGGAGCAACACCTGACCTTTCGGACATGCTTAAGT 600
DB 777 TCCTGCTGGGCGAGACACCCCGAGGAGCAACCAACCCCGACCTTTCAGATATGCTGAAT 836
QY 601 TTGAGAGTGACAAGCACACGAGACATCTCATGTGGAATATAGAGACACATCTCTCAACC 660
DB 837 TTGAGAGTGAGAAGCACACGAGACATCTTATGTGGAATACAGAGACATCTTCTCAACT 896
QY 661 TGTCCCTGAAGAAAGTGTCTTTCTTAGTGGGTGAGCACTTCTGCTCCAGAGCGCAGAGT 720
DB 897 TGTCTCGAAGAAAGTGTCTTTCTCAGTGGGTAACTACTTCTCTCCCGCAGACACTGAGT 956
QY 721 TTGCTCTCAAGGGCGATGACGCTGTTGTGGAACCCCATCATCATCTTAAITACTTGA 780
DB 957 TTGTTTCAAGGGCGATGACGATGTTTGTGTGGAACCCCATCATCATCTTAAITACTTGA 1016
QY 781 ATAGCTTATCCAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 840

	DB	1017	ATAGTTTATCCAAGACCAAAAGCATCTTCATAGTGATGTGATCCACAATGGCTG	1076
	QY	841	GGCTCACCGGGATAAAGAACACIAGAATACTACATCCCAGAAAGTCTTCTACACCAGGGCGTCT	903
	DB	1077	GACCTCATCGGATAAAGAAGCTGAAGTACTACATCCAGAAAGTTGTTACTCTGGCCTCT	1136
	QY	901	ACCCACCGTATGCGGGGGTGGTGATTCTGTACTCCGGGCCCTTGCCTTAGGGTGT	960
	DB	1137	ACCCACCCCTATGCAGGGGAGGGGGTTCCTCTACTCCGGCCACCTGGCCCCCTCAGGCTGT	1196
	QY	961	ACAGTGCACCTAGCGGGGTCCATCTCFACCCCTATTGATGATGTTTATACGGGAATGTGCC	1020
	DB	1197	ACCATATCACTGACCAAGTCCATCTCTACCCCATTTGNATGACGTTTTACTTGGAAATGTGCC	1256
	QY	1021	TTCAGAAACTGGGCGTTGTTCCAGAGAAGCACAAAGCGTTTCAGGACATTTTGAATAATGAAG	1080
	DB	1257	TTCAGAAACTCGGCGCTCGTTCCAGAGAAACACAAAGCGTTTCAGGACATTTTGAZATCCAGG	1316
	QY	1081	AGAAAAATAAGAAAATATTGTTTCCTTATACAGCCCTAATGTAGTACATAGCAGAGAAAC	1140
	DB	1327	AGAAAACAAAATTAACATCTGCTCCTCATGTAGATCTGATGTTAGTACATAGTAGAAAC	1376
	QY	1141	CTCAAGAGATGATTGAATATCTGGTCTCAGTTGCCAAAGTCCCTAATTTAAAATGGTGA	1236
	DB	1377	CTCAAGAGATGATTGATATTGGTCTTCAGTTGCGAGCTGCTCAATTTAAAATGGTAA	1432

RESULT 13

AC093401 99088 bp DNA linear PRI 16-APR-2002

LOCUS Homo sapiens BAC clone RP11-93M19 from 2, complete sequence.

DEFINITION

AC093401

ACCESSION

VERSION

KEYWORDS

HTG. 19807893

ORGANISM Homo sapiens (human)

SOURCE

REFERENCE

1 (bases 1 to 99088)

TITLE Toward a complete human genome sequence

JOURNAL Genome Res. 8 (11), 1097-1108 (1998)

MEDLINE 98063792

PUBMED 9847074

REFERENCE

2 (bases 1 to 99088)

AUTHORS Holmes,A. and Cotton,M.

TITLE The sequence of Homo sapiens BAC clone RP11-93M19

JOURNAL Unpublished (2001)

REFERENCE

3 (bases 1 to 99088)

AUTHORS Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (22-AUG-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE

4 (bases 1 to 99088)

AUTHORS Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (20-MAR-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE

5 (bases 1 to 99088)

AUTHORS Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (29-MAR-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE

6 (bases 1 to 99088)

AUTHORS Waterston,R.

TITLE Direct Submission

JOURNAL Submitted (16-APR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT

On Mar 29, 2002 this sequence version replaced gi:19424712.

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics -----
Center project name: H NH0093M19

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/qsc>

SOURCE INFORMATION:

The *SPC11* transposon-BAC library was made from the blood of one male donor. As described by Ozerow et al. (1998), the library was constructed by Taguchi, M., Carones, J.J., and de Jong, P.J. (1998). An improved approach for construction of bacterial artificial chromosome libraries Genomics 51:1-8. The clone may be obtained either from Research Genetics Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.choi1.org>
 VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-270B14, 2000 bp overlap;
the clone sequenced to the right is RP11-642D6, 2000 bp overlap.
Actual end of this clone is at base position 38298 of RP11-642D6.

Sequence derived from one M13 subclone, base position 30875 to 30886.

Polymorphisms have been identified between AC018462, AC093159 and AC093401.

FEATURES	source	Location/Qualifiers
		1. .99088
		/organism="Homo sapiens"
		/mol_type="genomic DNA"
		/db_xref="taxon:9606"
		/chromosome="2"
		/map="2"
		/clone="RP11-93M19"
		/Clone_lib="RPC1-11"
repeat_region		218. 337
		/rpt_family="MIR"
repeat_region		906. .1160
		/rpt_family="Alu"
repeat_region		1379. .1925
		/rpt_family="ERV1"
repeat_region		2261. .2297
		/rpt_family="MIR"
repeat_region		2571. .2939
		/rpt_family="L1"
repeat_region		2974. .3081
		/rpt_family="MIR"
repeat_region		3123. .3225
		/rpt_family="Alu"
repeat_region		3486. .3708
		/rpt_family="MIR"

```
misc_feature 3789..4320 /note="match to EST BG721682 (NID:g14000869)"
misc_feature 3829..4087 /note="match to EST AA377082 (NID:g2029630)"
misc_feature 3869..4691 /note="match to EST AU137528 (NID:g10998067)"
misc_feature 4100..4563 /note="match to EST H78875 (NID:g1056964) yu09g06.rl"
misc_feature 4141..4441 /note="match to EST AA263148 (NID:g:898954)"
misc_feature 4142..4793 /note="match to EST BG741357 (NID:g:4052010)"
misc_feature 4180..4678 /note="match to EST BG741357 (NID:g:4052010)"
misc_feature /note="similar to Mus musculus EST AA204363 (NID:g1801007) mu30ell1.rl"
misc_feature 4264..4724 /note="match to EST AW958548 (NID:g8148232)"
misc_feature 4426..5325 /note="similar to EST BM557603 (NID:g18799721)"
misc_feature 4468..4938 /note="similar to Mus musculus EST BE632125 (NID:g9914895) uu57d09.yl"
misc_feature 4498..4879 /note="match to EST H47990 (NID:g924042) yf75d11.rl"
misc_feature 4624..5231 /note="match to EST BF001654 (NID:g10701929)"
misc_feature 4666..5089 /note="match to EST AV682596 (NID:g10284459)"
misc_feature 4680..5391 /note="match to EST BE622082 (NID:g9893022)"
misc_feature 4915..5257 /note="match to EST R31722 (NID:g787565) yb63d09.rl"
misc_feature 4945..5390 /note="match to EST H13125 (NID:g877945) yj06e09.rl"
misc_feature 4948..5597 /note="match to EST BF131058 (NID:g10970087)"
misc_feature 5026 /note="match to EST R74552 (NID:g848922) y158a02.rl"
misc_feature 5031..5682 /note="match to EST BE564288 (NID:g9808008)"
misc_feature 5061..5553 /note="match to EST N58174 (NID:g1202064) yv66a07.sl"
misc_feature 5095..5554 /note="match to EST N66915 (NID:g1219040) za47c09.sl"
misc_feature 5107..5599 /note="match to EST R74552 (NID:g848922) y158a02.rl"
misc_feature 5128..5793 /note="match to EST BE865533 (NID:g10314309)"
misc_feature 5220..5836 /note="match to EST BF132275 (NID:g10971315)"
misc_feature 5254..5534 /note="match to EST BG319463 (NID:g13129329)"
misc_feature 5289..6180 /note="similar to EST BM550513 (NID:g18786776)"
misc_feature 5302..5473 /note="similar to Bos taurus EST BM087084 (NID:g16997712)"
misc_feature 5316..5838 /note="match to EST AA773646 (NID:g7825217) af75c02.rl"
misc_feature 5359..5986 /note="match to EST AV724225 (NID:g10828399)"
misc_feature 5482..6241 /note="match to EST BE738519 (NID:g10152511)"
misc_feature 5483..6278 /note="similar to Homo sapiens EST B5622910 (NID:g9893850)"
misc_feature 5503..5885 /note="similar to Bos taurus EST BM087084 (NID:g16997712)"
misc_feature 5641..6134 /note="match to EST AV715275 (NID:g10796792)"
misc_feature 5648..5885 /note="similar to Sus scrofa EST BF075149 (NID:g10868660)"
misc_feature 5682..6177 /note="match to EST AV705681 (NID:g10722976)"
```

```
misc_feature 5802..6310 /note="match to EST BF109367 (NID:g10939057)"
misc_feature 6009..6312 /note="match to EST AA835126 (NID:g2908854) ak63e07.sl"
misc_feature 6009..6311 /note="match to EST AV659072 (NID:g9880086)"
repeat_region 6010..6038 /rpt_family="AI-rich"
misc_feature 6224..8480 /note="match to EST D87736 (NID:g1552215)"
misc_feature 6229..6310 /note="match to EST D87736 (NID:g1552215)"

O-ery Match 75.8%; Score 906.4; DB 9; Length 99088;
Best Local Similarity 84.9%; Pred. No. 3.1e-244;
Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 1 AGATGAGTGTGGGGCTGCAAGAGTCAAGTTCCTGGGCATCCTGATGATGCAAAATGCT 60
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
3800 AAATGAGTGTGGAGCTGCAAGAAATAAGTTGTTGGGTATCCTGATGATGCAAAATGCT 3859
QY 61 TCATTATTGATGTGGAAGTCTCCAAAACAGTAGCCAAAGCAAAAATGGAAGGGAG 120
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
3860 TCATTATTATTTATGGAAGTCTCCAAAAGCAGTAGCCAAAGCAAAAATGGAAGGGG 3919
QY 121 GAGTAATAATCCGAAAGAGAGTCTGGAAGCCACCAGCACCTCCCGGGGCATCTGGA 180
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
3920 AAGTAATAATCCCAAAGAGAGTCTGGAAGATATCTACCCCTCCGAGGCATCTGGA 3979
QY 181 ACAGGGAACAGAGAGCTGAACAGGTGGTACAATGCCATCTTGAACAGGGTGGCCAACT 240
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
3980 ACCGAGAGCAAGAGAGCTGAACGGCAGTACAACCCCATCCTGAGCATGCTGACCAACC 4039
QY 241 AGACAGGGGAGCTAGCCACATCTCCAAACACAAGTCACTGAGCTATTGTGAACCACT 300
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
4040 AGACGGGGAGGGGGCAGGCTCTCCAATATGAGCCATCTGAACTACTGGAACCTGACC 4099
QY 301 CGACGGTCAACAGCTGTGACAGATTTAAATAATCTGCCGAGACAGATTTAAAGACTTTC 360
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
4100 TGAGGTCACGTCGGTGGTTACGGGTTTTAAACAATTCGCCGAGACATTTAAAGACTTTC 4159
QY 361 TCTTGTATTGAGATGCCGGAATTAATCGCTGCTATAGATCAACCGAAGAAATGICAA 420
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
4160 TGCTGTATTGAGATGCCGCAATTAATCACTGCTATAGATCAGCCGGATGAAGTGTGCA 4219
QY 421 AGAAGCCCTTCTTACTATTGGCGATAAGTCCCTCATTCACATTTGCCAGAGGCAAG 480
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
4220 AGAAACCTTTCTTGTGCTGGCGATTAAAGTCCCTCACTCCACATTTGCCAGAGGCAAG 4279
QY 481 CAATTGGGAGTCTTGGGGCGAGAACCAACGATAGGGAACACAGACAGTAGTAGGCTCT 540
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
4280 CAATCCGGGAATCTGGGGCCAAAGAACGACGAGGGAACCAACCGTGTGGGAGTCT 4339
QY 541 TCCCTGTGGGCAAGACACCCCGAGAGCAACCCCTGACCTTCGGACATGCTTAAGT 600
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
4340 TCCCTGTGGCCAGACACCCCGAGAGCAACCCCGACCTTTCAGATATGCTGAAT 4399
QY 601 TTGAGAGTGAACAGCAGGACATCTCATGTGGAACIATAGAGACATCTTCAACC 660
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
4400 TTGAGAGTGAAGCAACCAAGACATCTTATGTGAACIATAGAGACATCTTCTCAACT 4459
QY 661 TGTCTCTGAAGGAAGTGTGTTCTTAGTGGTGAGCACTTCTGTGTCAGACCCAGAGT 720
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
4460 TGTCTCTGAAGGAAGTGTGTTCTTAGTGGTGAGCACTTCTGTGTCAGACCCAGAGT 4519
QY 721 TTGCTTCAAGGGCGATGATGACGTGTTTGTGAACACCCATCATCTTAAATTAATTTGA 780
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
4520 TTGTTTCAAGGGCGATGACGATGTTTGTGAACACCCATCATCTTGAATTAATTTGA 4579
QY 781 ATAGCTTATCCAAAGAGCAAGCAAAAGACTTGTTCATAGGTGAGCTGATCCACAATGCT 840
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
4580 ATAGTTTATCCAAAGCAACCAAGCAAAAGACTTCTTTCATAGGTGATGATCCACAATGCT 4639
QY 841 GGCCTCACCGGATGAAGAACTGAAGTACTATACCCCAAGAGTCTTCTACACCCGGCTCT 900
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2003, 12:46:17 ; Search time 337.83 Seconds
(without alignments)
9556.669 Million coil updates/sec

Title: US-09-804-357B-1
 Perfect score: 1196
 Sequence: 1 aaatgaatgtggggctcga.....gtcctaattaaaactga 1196

Scoring table: IDENTITY NUC

Gapop 10.0 ; Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 51055512

Minimum DB seq length: 0

Maximum DB seq length:	200000000
Maximum DB seq length:	200000000

Post-processing: Minimum Match 08

Maximum Match 100%

Listing first 45 summaries

```

Database :
N_Geneseq_13Jun03.*
1: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSL/cgcdata/geneseq/geneseqn-emb1/NA2003.DAT.*

```

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1196	100.0	1196	21	AAA12714	CDNA encoding muri
2	1196	100.0	1196	23	AA167865	Murine LIG46 polyP
3	1196	100.0	1196	24	ABD46672	Murine LIG46 CDNA
4	1196	100.0	2876	21	AAZ88486	Murine Brainiac en
5	1196	100.0	2676	21	AAZ87185	Murine Brainiac CD
6	1194	99.8	1194	23	AA167866	Murine LIG46 polyP
7	1192.8	99.7	1260	25	ABV75081	Murine beta3Gnt ge
8	906.4	75.8	1707	21	AAA12716	CDNA encoding huma

ALIGNMENTS

RESULT 1
AAAl2714
ID AAAl2714 standard: cDNA; 1196 BP.

XX	AAA12714;
XX	
XX	25-JUL-2000 (first entry)
DT	
XX	cDNA encoding murine LIQ46 polypeptide.
DE	

xx Leptin; LIG46; body weight; leptin inducible gene; obesity; cachexia;
 KW LIG56; Tgtp; LRG-47; RC10-II; Stral3; ss.
 KW

XX	Mus sp.	Location/Qualifiers
OS	Key	3..1196
XX	CDS	/*tag= a
PH		/product= "LIG46"
FT		
FT		
FT		

XX
PN W0200015826-A2

XX
23-MAR-2000

XX
DE 10-SEP-1999. 99W0-11520722

XX
DD 10-SEP-1999. 08:15-0150957

PR 29-OCT-1998; 98US-0106378.
 10-NOV-1998; 98US-0106379.

PR 15-APR-1999; 99US-0292228.

PA (MILL-) MILLENNIUM PHARM INC.
XX White D, Zhou J, Tartaglia LA;
PI WPI: 2000-271461/23.
XX P-PSDB; AAY84641.
XX
PT Method for determining compounds which modulate body weight and can be
PT used to treat e.g. obesity comprises measuring the activity of leptin
PT Inducible genes -
XX
XX
XX Claim 16; Fig 1A-B; 123pp; English.
PS
XX
CC The present sequence encodes a leptin induced LIG46 polypeptide. The
CC specification describes a method for determining whether a compound can
CC be used to modulate body weight by measuring the activity of leptin
CC inducible genes, such as LIG46. The method can be used to specifically
CC detect LIG46 and LIG56 nucleic acid molecules relative to other nucleic
CC acid molecules encoding galactosyltransferases or GTP-binding proteins.
CC They can also be used in diagnostic assays to identify the presence or
CC absence of a genetic lesion or mutation characterized by aberrant
CC modification or mis-regulation of the genes or aberrant post-
CC translational modification of the proteins. LIG46 and LIG56 proteins
CC and nucleic acid antagonists can be used to treat obesity or cachexia.
CC LIG46 and LIG56 antagonists are used to treat obesity and LIG46 and
CC LIG56 agonists are used to treat low body weight. The leptin inducible
CC genes i.e. LIG46, LIG56, Tgtp, LRG-47, RC10-II and Stral3 can be used
CC to determine whether a compound modulates body weight and can then be
CC used to treat obesity or cachexia or low body weight.
XX
XX SQ Sequence 1196 BP; 350 A; 269 C; 290 G; 287 T; 0 other;
Query Match 100.0%; Score 1196; DB 21; Length 1196;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGATGAGTGTGGGCGTCGAAGAGTCAAGTTCGTGGGCATCTCTGATGATGGCAATGTCT 60
DB 1 AGATGAGTGTGGGCGTCGAAGAGTCAAGTTCGTGGGCATCTCTGATGATGGCAATGTCT 60
QY 61 TCATTTATTTGATTTGGAGTCTCCAAACACAGTAGCCCAAGACAAAATGGAAGGGAG 120
DB 61 TCATTTATTTGATTTGGAGTCTCCAAACACAGTAGCCCAAGACAAAATGGAAGGGAG 120
QY 121 GAGTAATAATCCCGAAGAGAGTCTCGAAGCACCAGCAGCTCCCGGGGATCTGGA 180
DB 121 GAGTAATAATCCCGAAGAGAGTCTCGAAGCACCAGCAGCTCCCGGGGATCTGGA 180
QY 181 ACAGGGAACAGAGAGAGTCTGACAGTGTGTAACATCCATCTGACAGAGGTCGCAATC 240
DB 181 ACAGGGAACAGAGAGAGTCTGACAGTGTGTAACATCCATCTGACAGAGGTCGCAATC 240
QY 241 AGACAGGGAGTCTGACAGATCTCCAAACACAGTCCAGTCTGACAGAGTCTGACAGAGT 300
DB 241 AGACAGGGAGTCTGACAGATCTCCAAACACAGTCCAGTCTGACAGAGTCTGACAGAGT 300
QY 301 CGACGGTCTATGACAGTCTGACAGATTTTAATATCTGCGGACAGATTTAAGAGTCTTC 360
DB 301 CGACGGTCTATGACAGTCTGACAGATTTTAATATCTGCGGACAGATTTAAGAGTCTTC 360
QY 361 TCTGTATTTGATGATCGCGAATTTACTCGTCTTATAGATCAACCGAAGAAATGTGCAA 420
DB 361 TCTGTATTTGATGATCGCGAATTTACTCGTCTTATAGATCAACCGAAGAAATGTGCAA 420
QY 421 AGAAGCCCTCTTACTATTGGCGATAAGTCCCTCATCTCCACATTTGGCAGAGGCAAG 480
DB 421 AGAAGCCCTCTTACTATTGGCGATAAGTCCCTCATCTCCACATTTGGCAGAGGCAAG 480
QY 481 CAATTCGGAGTCTTGGGCGCGAAGAACCAACAGTACAGGAGTGTAGGAGTCT 540
DB 481 CAATTCGGAGTCTTGGGCGCGAAGAACCAACAGTACAGGAGTGTAGGAGTCT 540
QY 541 TCCTGTTGGGCAAGACACCCCGAGAGGACAAACCCCTGACCTTTTCGGACATGCTTAAGT 600

DB 541 TCCTGTTGGGCAAGACACCCCGAGAGGACAAACCCCTGACCTTTTCGGACATGCTTAAGT 600
QY 601 TTGAGAGTGAACAGCACCAGGAGATCTCTCATGTGGAACTATAGAGACATTTTCAACC 660
DB 601 TTGAGAGTGAACAGCACCAGGAGATCTCTCATGTGGAACTATAGAGACATTTTCAACC 660
QY 661 TGTCCTTGAAGGAGTGTCTTTCTTAGTGGGTGAGCAGTCTCTCTCCAGAGCGAGAGT 720
DB 661 TGTCCTTGAAGGAGTGTCTTTCTTAGTGGGTGAGCAGTCTCTCTCCAGAGCGAGAGT 720
QY 721 TTGCTTCAAGGGGATGATGAGTGTCTTGAACACCCATCATCTCTTAATTAATCTGA 780
DB 721 TTGCTTCAAGGGGATGATGAGTGTCTTGAACACCCATCATCTCTTAATTAATCTGA 780
QY 781 ATAGCTTTATCCAAAGAGCAAAAGCAAGTGTTCATAGGTGACGTGATCCCAATGCTG 840
DB 781 ATAGCTTTATCCAAAGAGCAAAAGCAAGTGTTCATAGGTGACGTGATCCCAATGCTG 840
QY 841 GGCCTCACCGGGATAGAAACTGAACTACTACATCCAGAGTCTTCTACACCGGCTCT 900
DB 841 GGCCTCACCGGGATAGAAACTGAACTACTACATCCAGAGTCTTCTACACCGGCTCT 900
QY 901 ACCCACCGTATCCCGGGGTGGTGGATTCTCTACTCCGGCCCCCTTGGCTTGAGGCTGT 960
DB 901 ACCCACCGTATCCCGGGGTGGTGGATTCTCTACTCCGGCCCCCTTGGCTTGAGGCTGT 960
QY 961 ACAGTGGAGTACCGGGGTCATCTCTACCTATTGATGATGATTTACGGGAAATGTGC 1020
DB 961 ACAGTGGAGTACCGGGGTCATCTCTACCTATTGATGATGATTTACGGGAAATGTGC 1020
QY 1021 TTCAGAACTGGGCTTGTTCAGAGAGCAAAAGGCTTCAGAGATTTGATTTGAAG 1080
DB 1021 TTCAGAACTGGGCTTGTTCAGAGAGCAAAAGGCTTCAGAGATTTGATTTGAAG 1080
QY 1081 AGAAAAATAGAAAAATATTTTGTTCATATAGACCTAATGTTAGTACATAGCAGAAAC 1140
DB 1081 AGAAAAATAGAAAAATATTTTGTTCATATAGACCTAATGTTAGTACATAGCAGAAAC 1140
QY 1141 CTCAGAGATGATGATGATCTGCTCAGTTGCAAGTCTCTAATTTAAAAATGCTGA 1196
DB 1141 CTCAGAGATGATGATGATCTGCTCAGTTGCAAGTCTCTAATTTAAAAATGCTGA 1196
RESULT 2
AA167865
ID AA167865 standard; cDNA; 1196 BP.
XX
AC AA167865;
XX
DT 13-MAR-2002 (first entry)
XX
DE Murine LIG46 polypeptide encoding cDNA.
XX
KW Leptin; LIG46; LIG56; leptin induced gene; Tgtp; LRG-47; RC10-II; Stral3;
KW anorectic; anabolic; antisense therapy; mouse; ss.
XX
OS Mus sp.
XX
EH Key Location/Qualifiers
FT CDS 3..1194
FT /tag= a
FT /product= "LIG46"
FT /note= "the coding sequence is specifically claimed"
FT sig_peptide 3..98
FT /tag= b
FT mat_peptide 99..1193
FT /tag= c
XX
XX US2001024808-A1.
XX
XX 27-SEP-2001.
XX

PF 12-MAR-2001; 2001US-0804357.
XX
PR 29-OCT-1998; 98US-106378P.
PR 19-NOV-1998; 98US-0195896.
PR 10-SEP-1998; 98US-0150857.
XX
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI White D, Zhou J, Tartaglia LA;
XX WPI: 2001-624963/72.
XX P-PSDB; AAG66115, AAG66116.
DR
XX An isolated polypeptide useful for treating a weight disorder and for
PT screening compounds that may inhibit protein activity comprises a
PT leptin-induced protein
XX
XX Claim 16; Fig 1: 46pp; English.
XX
CC The invention relates to genes whose expression are induced by leptin.
CC Lig46 and Lig56 are novel leptin induced genes (lig), while four other
CC genes such as Tgtp, LR4-47, RL10-11 and Stral3 have been previously
CC identified. The leptin induced proteins can be expressed by standard
CC recombinant methodology. The proteins and encoding polynucleotides may
CC be used in screening assays to identify compounds that may bind to it.
CC Administering a molecule, e.g., an antisense molecule, which reduces
CC expression of activity of protein selected from Lig46, Lig56, Tgtp,
CC LR4-47, RL10-11, and Stral3 may be used to treat a weight disorder. The
CC method may further comprise administering leptin. The present sequence
CC represents a cDNA encoding a murine Lig46 protein.
XX
XX Sequence 1196 BP; 350 A; 269 C; 290 G; 287 T; 0 other;
XX

Query Match 100.0%; Score 1196; DB 23; Length 1196;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGATGAGTGTGGGCGTGGAGAGTCAAGTGTCTGGGATCTCTGATGGCAATGTCT 60
DB |||||||
QY 1 AGATGAGTGTGGGCGTGGAGAGTCAAGTGTCTGGGATCTCTGATGGCAATGTCT 60
DB |||||||
QY 61 TCATTTATTTGATTTGGAAAGTCTCCAAACAGTACCCCAAGACAAAATGGAAGGGAG 120
DB |||||||
QY 61 TCATTTATTTGATTTGGAAAGTCTCCAAACAGTACCCCAAGACAAAATGGAAGGGAG 120
DB |||||||
QY 121 GAGTAATATCCGGAAGAGAGTCTTGGAGCCACCCAGCCTCCCGGGCATCTGGA 180
DB |||||||
QY 121 GAGTAATATCCGGAAGAGAGTCTTGGAGCCACCCAGCCTCCCGGGCATCTGGA 180
DB |||||||
QY 181 ACAGGGAACAGGAGAGTCAACAGGTGTGACAAATCCCATCTTGAACAGGGTGGCAATC 240
DB |||||||
QY 181 ACAGGGAACAGGAGAGTCAACAGGTGTGACAAATCCCATCTTGAACAGGGTGGCAATC 240
DB |||||||
QY 241 AGACAGGGAGCTAGGCACATCTCCAAACAGTACACCTGAGCTATTGTGAACAGACT 300
DB |||||||
QY 241 AGACAGGGAGCTAGGCACATCTCCAAACAGTACACCTGAGCTATTGTGAACAGACT 300
DB |||||||
QY 301 CGACGGTCATGACAGCTGTACAGATTTTAATATCTGCGGACAGNTTAAAGACTTTC 360
DB |||||||
QY 301 CGACGGTCATGACAGCTGTACAGATTTTAATATCTGCGGACAGNTTAAAGACTTTC 360
DB |||||||
QY 361 TCTTGATTTGAGATCGCGGAATTAATCTGCTGCTTATAGATCAACCGAAGAAATGTGCAA 420
DB |||||||
QY 361 TCTTGATTTGAGATCGCGGAATTAATCTGCTGCTTATAGATCAACCGAAGAAATGTGCAA 420
DB |||||||
QY 421 AGAAGCCCTTCTTACTATTGGCGATAAGTCCCTCATTCACATTTTGCAGAGGCAAG 480
DB |||||||
QY 421 AGAAGCCCTTCTTACTATTGGCGATAAGTCCCTCATTCACATTTTGCAGAGGCAAG 480
DB |||||||
QY 481 CAATTCGGGAGTCTTGGGCGCGAGAACCAACGTAGGGACACAGACAGTAGTGGGTCT 540
DB |||||||
QY 481 CAATTCGGGAGTCTTGGGCGCGAGAACCAACGTAGGGACACAGACAGTAGTGGGTCT 540
DB |||||||

QY 541 TCCTGTGGCAAGACACACCCCGAGGAGACACACCCTTGACCTTTCGGACATGCTTAAGT 600
DB |||||||
QY 541 TCCTGTGGCAAGACACACCCCGAGGAGACACACCCTTGACCTTTCGGACATGCTTAAGT 600
DB |||||||
QY 601 TTGAGAGTGACAAGCACCAGGACATCCTCATGTGGAACTATAGAGACACATCTTCAACC 660
DB |||||||
QY 601 TTGAGAGTGACAAGCACCAGGACATCCTCATGTGGAACTATAGAGACACATCTTCAACC 660
DB |||||||
QY 661 TGTCCCTGAAGGAAGTGTCTTTTCTTAGTGGGTGAGCATTCTCTCCAGACACAGAGT 720
DB |||||||
QY 661 TGTCCCTGAAGGAAGTGTCTTTTCTTAGTGGGTGAGCATTCTCTCCAGACACAGAGT 720
DB |||||||
QY 721 TTGTCITCAAGGGCGATGATGACGTGTTTGTGAACACCCATCATCTTAAATCTTGA 780
DB |||||||
QY 721 TTGTCITCAAGGGCGATGATGACGTGTTTGTGAACACCCATCATCTTAAATCTTGA 780
DB |||||||
QY 781 ATAGCTTATCCAAGAGCAAAAGCACTTGTTCATAGTGAGTGATCCACAAATGCTG 840
DB |||||||
QY 781 ATAGCTTATCCAAGAGCAAAAGCACTTGTTCATAGTGAGTGATCCACAAATGCTG 840
DB |||||||
QY 841 GGCCTCACCGGATGAAGAACTGAAGTACTACATCCCAAGAGTCTTCTACACGGCGTCT 900
DB |||||||
QY 841 GGCCTCACCGGATGAAGAACTGAAGTACTACATCCCAAGAGTCTTCTACACGGCGTCT 900
DB |||||||
QY 901 ACCCACCCTATGCCGGGGTGGTGATTCCTGTACTCCGGCCCTTGCCTTGAGGCTGT 960
DB |||||||
QY 901 ACCCACCCTATGCCGGGGTGGTGATTCCTGTACTCCGGCCCTTGCCTTGAGGCTGT 960
DB |||||||
QY 961 ACAGTGGCACTAGCGGGTCCATCTCTACCTATTGATGATGTTTATACGGGAATGTGCC 1020
DB |||||||
QY 961 ACAGTGGCACTAGCGGGTCCATCTCTACCTATTGATGATGTTTATACGGGAATGTGCC 1020
DB |||||||
QY 1021 TTCAGAAACTGGCCCTTCTCCAGAGAGCAAAAGGCTTCAGGACATTTGATATTGAG 1080
DB |||||||
QY 1021 TTCAGAAACTGGCCCTTCTCCAGAGAGCAAAAGGCTTCAGGACATTTGATATTGAG 1080
DB |||||||
QY 1081 AGAAAAATAAGAAAAATATTGTTCCTATATAGACCTAATGTAGTACATAGCAAAAC 1140
DB |||||||
QY 1081 AGAAAAATAAGAAAAATATTGTTCCTATATAGACCTAATGTAGTACATAGCAAAAC 1140
DB |||||||
QY 1141 CTCAGAGATGATTGATATCTGGTCTCAGTTCGCAAGTCCCTAAATTTAAATGCTGA 1196
DB |||||||
QY 1141 CTCAGAGATGATTGATATCTGGTCTCAGTTCGCAAGTCCCTAAATTTAAATGCTGA 1196
DB |||||||

RESULI 3
AAD46672
ID AAD46672 standard; cDNA; 1196 BP.
XX
AC AAD46672;
XX
DT 27-JAN-2003 (first entry)
XX
DE Murine Lig46 cDNA.
XX
KW Lig46; leptin; low body weight; chromosomal mapping; tissue typing;
KW forensic biology; transgenic; gene therapy; antianorectic; murine;
KW gene; ss.
XX
OS Mus sp.
XX
XX Key Location/Qualifiers
FH 3..1196
CDS /tag- a
FT /product= "Murine Lig46 protein"
FT 3..98
FT sig_peptide /tag- b
FT 99..1193
FT mat_peptide /tag- c
FT /product= "Murine mature Lig46 protein"
XX
PN W0200274905-A2.
XX

XX WP1: 2000-161481/15.
 DR P-PSDB: AAY79953.
 XX
 XX Mammalian EGGHEAD and BRAINAC proteins which mediate cell to cell
 PT adhesion and may be used to treat cancer, psoriasis and other skin
 PT lesions and nervous system defects or diseases -
 XX
 PS Example 1: Page 24-25; 30pp: English.
 XX
 CC The present invention describes mammalian EGGHEAD and BRAINAC proteins,
 CC which mediate cell-to-cell adhesion. The mammalian proteins are similar
 CC to the EGGHEAD and BRAINAC genes of Drosophila melanogaster. BRAINAC
 CC and EGGHEAD genes regulate adhesion between epithelial cells and require
 CC the presence of the Notch protein for this activity. Mutational
 CC disruption of BRAINAC, EGGHEAD or Notch results in the loss of follicle
 CC epithelial cell adhesion, thereby reducing the efficiency of signalling
 CC through other epithelial cell receptors such as the epidermal growth
 CC factor receptor. The BRAINAC and EGGHEAD proteins have cytosolic,
 CC dermatological and neuroactive properties. As the BRAINAC and EGGHEAD
 CC proteins are involved in adhesion between epithelial cells, and as this
 CC adhesion is required for cell viability, cell growth regulation and cell
 CC fate specification, it is envisioned that wild type or mutant forms of
 CC mammalian BRAINAC and/or EGGHEAD can be used to alter epithelial cell
 CC adhesion. This should be useful in treating many diseases which present
 CC problems of cell viability, cell growth regulation and cell fate
 CC specification. For example, these proteins, or active fragments or
 CC analogues of these proteins and these genes can be used to treat
 CC diseases such as cancer, psoriasis and other skin lesions, and nervous
 CC system defects or diseases. The present sequence encodes the murine
 CC BRAINAC protein.
 XX
 SQ Sequence 2676 BP; 732 A; 562 C; 637 G; 744 T; 1 other;

Query Match 100.08; Score 1196; DB 21; Length 2676;

Best Local Similarity 100.08; Pred. No. 3;
 Matches 1196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGATGAGTGTGGGGCTCGAAGAGTCAAGTTCGCTGGGCATCCTGATGATGCAATGCT 60
 DB 208 AGATGAGTGTGGGGCTCGAAGAGTCAAGTTCGCTGGGCATCCTGATGATGCAATGCT 267
 QY 61 TCATTATTGTTGGAAGTCTCCAAAACAGTAGCCAGAGCAAAAATGGAAAGGAG 120
 DB 268 TCATTATTGTTGGAAGTCTCCAAAACAGTAGCCAGAGCAAAAATGGAAAGGAG 327
 QY 121 GAGTAATAATCCGAAACAGAGTCTCGAAGCCACCCAGCACTCCCGGCATCTGGA 180
 DB 328 GAGTAATAATCCGAAACAGAGTCTCGAAGCCACCCAGCACTCCCGGCATCTGGA 387
 QY 181 ACAGGGAACAGAGAGTGTGAACAGTGTGATACATCCCATCTTGAACAGGGTGGCCATC 240
 DB 388 ACAGGGAACAGAGAGTGTGAACAGTGTGATACATCCCATCTTGAACAGGGTGGCCATC 447
 QY 241 ACAGAGGGGAGCTAGCCACATCTCCAAACACAAGTCACTGAGCTATTGTGAACAGACT 300
 DB 448 ACAGAGGGGAGCTAGCCACATCTCCAAACACAAGTCACTGAGCTATTGTGAACAGACT 507
 QY 301 CCACGGTCTATGACGCTGTGACAGATTTTAATAATCTGCGGACAGATTTAAAGACTTTC 360
 DB 508 CGACGGTCTATGACGCTGTGACAGATTTTAATAATCTGCGGACAGATTTAAAGACTTTC 567
 QY 361 TCTTGTATTGAGTCCGGATTAATCTCGCTGCTTTATGATCAACCCAGAAATGTGCAA 420
 DB 568 TCTTGTATTGAGTCCGGATTAATCTCGCTGCTTTATGATCAACCCAGAAATGTGCAA 627
 QY 421 AGAAGCCCTTCTTACTATTGCGGATTAAGTCCCTCATCTCCACATTTGCCAGAGGCAAG 480
 DB 628 AGAAGCCCTTCTTACTATTGCGGATTAAGTCCCTCATCTCCACATTTGCCAGAGGCAAG 687
 QY 481 CAAATTCGGGAGTCTTGGGGCCGAGAAACCAACGCTAGGGAACAGACAGTACTGAGGTCT 540
 DB 688 CAAATTCGGGAGTCTTGGGGCCGAGAAACCAACGCTAGGGAACAGACAGTACTGAGGTCT 747

QY 541 TCCTGTGGGCAAGACACCCCGAGAGCAACACCCCTGACCTTCCTTCGACATGCTTAAGT 600
 DB 748 TCCTGTGGGCAAGACACCCCGAGAGCAACACCCCTGACCTTCCTTCGACATGCTTAAGT 807
 QY 601 TTGAGAGTGAACAGACACCCAGAGCATCTCTCATGTGGAACATATAGAGACACATTTCTCAACC 660
 DB 808 TTGAGAGTGAACAGACACCCAGAGCATCTCTCATGTGGAACATATAGAGACACATTTCTCAACC 867
 QY 661 TGTCCCTCAAGGAAGTGTCTTTCTTAGGTGGGTGAGCAGCTTCTCTCCAGAGCGCAGAGT 720
 DB 868 TGTCCCTCAAGGAAGTGTCTTTCTTAGGTGGGTGAGCAGCTTCTCTCCAGAGCGCAGAGT 927
 QY 721 TTGCTTCAAGGGGATGATGACGTGTTTGTGAACCCATACATCCTTAATTAATCTGA 780
 DB 928 TTGCTTCAAGGGGATGATGACGTGTTTGTGAACCCATACATCCTTAATTAATCTGA 987
 QY 781 ATAGCTTATCCAAGAGCAAAAGCAAGCTTGTTCATAGGTGAGTGTGATCCACAATGCTG 840
 DB 988 ATAGCTTATCCAAGAGCAAAAGCAAGCTTGTTCATAGGTGAGTGTGATCCACAATGCTG 1047
 QY 841 GGCCTCACGGGATGAAGACTGAAGTACTACATCCAGAAAGTCTTCTACACGGCGCTCT 900
 DB 1048 GGCCTCACGGGATGAAGACTGAAGTACTACATCCAGAAAGTCTTCTACACGGCGCTCT 1107
 QY 901 ACCCACCGTATCGCGGGGTGGTGGATTCTCTGATCTCGGCCCTTGCCTTGAGGCTGT 960
 DB 1108 ACCCACCGTATCGCGGGGTGGTGGATTCTCTGATCTCGGCCCTTGCCTTGAGGCTGT 1167
 QY 961 ACAGTGCAGTACGCGGTGCTCATCTCTACCTATTGATGATGTTTATACGGGAATGTGCC 1020
 DB 1168 ACAGTGCAGTACGCGGTGCTCATCTCTACCTATTGATGATGTTTATACGGGAATGTGCC 1227
 QY 1021 TTCAGAAATCGGCTTCTCCAGAGCAACAGGCTTCAGGACATTTGATATTGAAG 1080
 DB 1228 TTCAGAAATCGGCTTCTCCAGAGCAACAGGCTTCAGGACATTTGATATTGAAG 1287
 QY 1081 AGAAAAATGAAGAAAAATATTTTCTCTATATAGACCTAATGTTAGTACATAGCAGAAAAAC 1140
 DB 1288 AGAAAAATGAAGAAAAATATTTTCTCTATATAGACCTAATGTTAGTACATAGCAGAAAAAC 1347
 QY 1141 CTCAGAGATGATGATATCTGCTCAGTTCAGTTCGAAAGTCTTAATTTAAATGCTGA 1196
 DB 1348 CTCAGAGATGATGATATCTGCTCAGTTCGAAAGTCTTAATTTAAATGCTGA 1403
 RESULT 5
 ID AA287185 standard; cDNA; 2676 BP.
 AC AA287185;
 DT 08-MAY-2000 (first entry)
 XX Murine Brainiac cDNA.
 XX Brainiac; murine; mammalian; expressed sequence tag; EST;
 KW Drosophila melanogaster; Egghead; Notch; epithelial cell adhesion;
 KW Fringe family; cell viability; growth regulation; cell fate; cancer;
 KW psoriasis; skin lesion; nervous system disorder; developmental syndrome;
 KW transgenic animal; drug screening; ss.
 XX Mus sp.
 OS
 XX Key Location/Qualifiers
 FH 210..1403
 ET CDS /tag- a
 FT /product= "Murine Brainiac protein"
 XX CA2255109-A1.
 XX
 PD 17-JUN-1999.
 XX

PF 17-DEC-1998; 98CA-2255109.
XX
PR 17-DEC-1997; 97CA-2225126.
XX
PA (HSCR-) HSC RES & DEV LP.
XX
PI Ega: SE;
XX
DR WPI: 2000-148082/14.
DR P-PSDB; AAY69697.
XX
PT New nucleic acids encoding a murine and human Brainiac protein. The cDNA
PT for detecting somatic or germline DNA lesions which are responsible for
PT developmental syndromes or diseases including cancer -
XX
PS Claim 5; Page 26-27; 40pp; English.
XX
CC This sequence represents cDNA encoding murine Brainiac protein. The cDNA
CC was isolated from a mouse mammary gland cDNA library via the use of a
CC probe generated via PCR from a variety of murine tissues. The PCR
CC primers used for probe generation (AAZB7187-287188) were based on EST
CC (expressed sequence tag) sequences with significant similarity to
CC Drosophila melanogaster Brainiac. Drosophila Brainiac and Egghead
CC proteins regulate adhesion between epithelial cells, this activity being
CC dependent on the presence of Notch. Drosophila Brainiac is a secreted
CC protein, and has sequence similarities with the Drosophila Fringe
CC proteins. Brainiac/ Egghead-mediated epithelial cell adhesion is required
CC for cell viability, cell growth regulation and cell fate specification.
CC Wild-type or mutant forms of mammalian Brainiac proteins may therefore be
CC used to alter epithelial cell adhesion in a mammal. Mammalian Brainiac
CC proteins, active fragments analogues, and nucleic acids may be used to
CC treat diseases such as cancer, psoriasis and other skin lesions, and
CC nervous system disorders. Mammalian Brainiac nucleic acids may also be
CC used to detect somatic or germline DNA lesions which are responsible for
CC developmental syndromes or diseases including cancer. The mammalian
CC Brainiac proteins and fragments or its analogues are useful as antigens
CC in immunoassays including enzyme-linked immunosorbent assays (ELISA),
CC radioimmunoassays (RIA) and other non-enzyme linked antibody binding
CC assays. Non-human transgenic animals comprising nucleotide sequences
CC encoding human Brainiac protein (AAY69698) can be used as animal models
CC for the study of mammalian Brainiac gene function, for the screening of
CC candidate compounds and for the evaluation of potential therapeutic
CC interventions.
XX
SQ Sequence 2676 BP; 732 A; 562 C; 637 G; 744 T; 1 other;
Query Match 100.0%; Score 1196; DB 21; Length 2676;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGATCAGTGTGGGGTCGAGAGTCAAGTGTCTGGGCATCCTGTATGATGGCAAAATGTCT 60
DB 208 AGATCAGTGTGGGGTCGAGAGTCAAGTGTCTGGGCATCCTGTATGATGGCAAAATGTCT 267
QY 61 TCATTTATTTGATTGTGGAAGTCTCCAAAACAGTAGCCAGACAAAATGGAAAGGGAG 120
DB 268 TCATTTATTTGATTGTGGAAGTCTCCAAAACAGTAGCCAGACAAAATGGAAAGGGAG 327
QY 121 GAGTAAATATCCGAAAGAGAAGTTCTGGAAGCCACCAGCACTCCCGGGCATATCTGGA 180
DB 328 GAGTAAATATCCGAAAGAGAAGTTCTGGAAGCCACCAGCACTCCCGGGCATATCTGGA 387
QY 181 ACAGGGAACAGGAGAGTGAACAGTGGTACATCCATCTTGACAGGGTGGCCAAATC 240
DB 388 ACAGGGAACAGGAGAGTGAACAGTGGTACATCCATCTTGACAGGGTGGCCAAATC 447
QY 241 AGACAGGGGAGTACGCCATCTCCAAACAGTACACCTGAGCTATTTGTGAACACAGACT 300
DB 448 AGACAGGGGAGTACGCCATCTCCAAACAGTACACCTGAGCTATTTGTGAACACAGACT 507
QY 301 CGACGGTCATCACAGCTGTGACAGATTTTAATAATCTGCCGGGACAGATTTAAAGACTTTC 360
DB 508 CGACGGTCATCACAGCTGTGACAGATTTTAATAATCTGCCGGGACAGATTTAAAGACTTTC 567

QY 361 TCTGTATTGAGATGCCGAATTACTCGCTGCTTATAGATCAACCGAAGAAATGTGCA 420
DB 568 TCTGTATTGAGATGCCGAATTACTCGCTGCTTATAGATCAACCGAAGAAATGTGCA 627
QY 421 AGAAGCCCTTCTTACTATTGGCGATAAGTCCCTCAITCCACATTTTGCACGAAGCGAAG 480
DB 628 AGAAGCCCTTCTTACTATTGGCGATAAGTCCCTCAITCCACATTTTGCACGAAGCGAAG 587
QY 481 CAATTCGGGAGTCTTGGGGCCGAGAAACCAACCTAGGGAACCCAGACAGTAGTGAGGGTCT 540
DB 688 CAATTCGGGAGTCTTGGGGCCGAGAAACCAACCTAGGGAACCCAGACAGTAGTGAGGGTCT 747
QY 541 TCCTGTGGGACAGACACCCCCAGAGACAAACCAACCCCTGACCTTTCGGACATGCTTAAGT 600
DB 748 TCCTGTGGGACAGACACCCCCAGAGACAAACCAACCCCTGACCTTTCGGACATGCTTAAGT 807
QY 601 TTGAGGTGACAGACACCCAGAGACATCCTCATGTGGAACTATAGAGACACATTTCTCAACC 660
DB 808 TTGAGGTGACAGACACCCAGAGACATCCTCATGTGGAACTATAGAGACACATTTCTCAACC 867
QY 661 TGTCCCTGAAAGAGTGTCTTTCTTGTAGTGGGTGAGCACITTCCTGTCCAGACGAGAGT 720
DB 868 TGTCCCTGAAAGAGTGTCTTTCTTGTAGTGGGTGAGCACITTCCTGTCCAGACGAGAGT 927
QY 721 TTGTCTTCAAGGGCGATGATGAGTGTCTTGTGAACCCCATCACATCCTTAATCTTGA 780
DB 928 TTGTCTTCAAGGGCGATGATGAGTGTCTTGTGAACCCCATCACATCCTTAATCTTGA 987
QY 781 ATAGCTTATCCAAAGACAAAGCCAAAGACTTGTTCATAGTGACGTGATCCCAATGTCTG 840
DB 988 ATAGCTTATCCAAAGACAAAGCCAAAGACTTGTTCATAGTGACGTGATCCCAATGTCTG 1047
QY 841 GGCCTCACCGGATGAAGTGAAGTACTACATCCCAAGAGTCTTCTACACCGGCTCT 900
DB 1048 GGCCTCACCGGATGAAGTGAAGTACTACATCCCAAGAGTCTTCTACACCGGCTCT 1107
QY 901 ACCCAGCTATGCGGGGGTGGTGGATTCCTGACTCCGGCCCTTGGCTTGAGGCTGT 960
DB 1108 ACCCAGCTATGCGGGGGTGGTGGATTCCTGACTCCGGCCCTTGGCTTGAGGCTGT 1167
QY 961 ACAGTGGGACTAGCGGGTCCATCTTACCTTATGATGATGTTTATAGGGGAATGTCC 1020
DB 1168 ACAGTGGGACTAGCGGGTCCATCTTACCTTATGATGATGTTTATAGGGGAATGTCC 1227
QY 1021 TTCAGAAACTGGGCTTGTTCAGAGAAAGCACAAGAGCTTCAGACATTTGATATTGAAG 1080
DB 1228 TTCAGAAACTGGGCTTGTTCAGAGAAAGCACAAGAGCTTCAGACATTTGATATTGAAG 1287
QY 1081 AGAAATAAGAAATAATTTGCTTATATAGACCTAATGTAGTACATAGCAGAAAC 1140
DB 1288 AGAAATAAGAAATAATTTGCTTATATAGACCTAATGTAGTACATAGCAGAAAC 1347
QY 1141 CTCAGAGATGATTGATATCTGGTCTCAGTGTGCAAGTCTCTAATTTTAAATGCTGA 1196
DB 1348 CTCAGAGATGATTGATATCTGGTCTCAGTGTGCAAGTCTCTAATTTTAAATGCTGA 1403

RESULT 6
AA167866
ID AA167866 standard; cDNA; 1194 BP.
XX
AC AA-57866;
XX
DT 13-MAR-2002 (first entry)
XX
DE Murine IIG46 polypeptide coding sequence.
XX
KW Leptin; LIG46; LIG56; leptin induced gene; Tgtp; LRG-47; RC10-II; Stral3;
KW anorectic; anabolic; antisense therapy; mouse; ss.
XX
OS Mus sp.
XX

FH Key Location/Qualifiers
 FT CDS 1..1194
 FT /*tag= a
 FT /product= "LIG46"
 FT sig_peptide 1..96
 FT mat_peptide 97..1191
 FT /*tag= b
 FT /*tag= c
 XX
 XX US2001024808-A1.
 PN 27-SEP-2001.
 XX
 XX 12-MAR-2001; 2001US-0804357.
 PR 29-OCT-1998; 98US-106378P.
 PR 19-NOV-1998; 98US-0195896.
 PR 10-SEP-1998; 98US-0150857.
 XX
 XX (MILL-) MILLENNIUM PHARM INC.
 PA White D, Zhou J, Tartaglia LA;
 PI WPI: 2001-624963/72
 DR P-PSDB; AAG66115, AAG66116.
 DR
 XX An isolated polypeptide useful for treating a weight disorder and for
 PT screening compounds that may inhibit protein activity comprises a
 PT lepton-induced protein -
 XX
 XX Claim 16; Fig 1; 46pp; English.
 XX
 CC The invention relates to genes whose expression are induced by leptin.
 CC LIG46 and LIG56 are novel leptin induced genes (LIG), while four other
 CC genes such as Tgtp, LRG-47, RC10-1f and Stral3 have been previously
 CC identified. The leptin induced proteins can be expressed by standard
 CC recombinant methodology. The proteins and encoding polynucleotides may
 CC be used in screening assays to identify compounds that may bind to it.
 CC Administering a molecule, e.g., an antisense molecule, which reduces
 CC expression of activity of protein selected from LIG46, LIG56, Tgtp,
 CC LRP-47, RC10-1f, and Stral3 may be used to treat a weight disorder. The
 CC method may further comprise administering leptin. The present sequence
 CC represents the murine LIG46 protein coding sequence.
 XX
 XX Sequence 1194 BP; 349 A; 269 C; 289 G; 287 T; 0 other;
 SQ

Query Match 99.8%; Score 1194; DB 23; Length 1194;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ATGAGTGTGGGGCTCGAAGAGTCAAGTTCGTGGGATCCTGATGATGCAATGCTTC 62
 DB 1 ATGAGTGTGGGGCTCGAAGAGTCAAGTTCGTGGGATCCTGATGATGCAATGCTTC 60
 QY 63 ATTTATTTGATTCGTGAAGTCTCCAAAACAGTAGCCAAAGAAAATGGAAGGAGGA 122
 DB 61 ATTTATTTGATTCGTGAAGTCTCCAAAACAGTAGCCAAAGAAAATGGAAGGAGGA 120
 QY 123 GTAATATCCGAAAGAGAGTTCCTGGAAGCCACCCAGCAGCTCCCGGGGCATCTGGAA 182
 DB 121 GTAATATCCGAAAGAGAGTTCCTGGAAGCCACCCAGCAGCTCCCGGGGCATCTGGAA 180
 QY 183 AGGACACAGAGAGTGTACAGGTGGTACATCCCATCTTGAACAGGTGGCCCAATCAG 242
 DB 181 AGGAAACAGGAGAGTGTACAGGTGGTACATCCCATCTTGAACAGGTGGCCCAATCAG 240
 QY 243 ACAGGGGAGCTAGCCATCTCCAAACAGAGTCACTGAGCTATTTGTGAACAGGACTCG 302
 DB 241 ACAGGGGAGCTAGCCATCTCCAAACAGAGTCACTGAGCTATTTGTGAACAGGACTCG 300
 QY 303 ACGGTCATGACAGCTGTGACAGATTTTAATAATCTGCCGGACAGATTTAAAGACTTTCTC 362
 DB

Db 301 ACGTCATGACAGCTGTGACAGATTTTAATATCTGCCGACAGATTAAAGACTTCTC 360
 QY 363 TTGATTTGAGATGCCGGAATTACTCGCTCTTATAGATCAACCCGAGAAATGIGCAAG 422
 Db 361 TTGATTTGAGATGCCGGAATTACTCGCTCTTATAGATCAACCCGAGAAATGIGCAAG 420
 QY 423 AAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAGCA 482
 Db 421 AAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAGCA 480
 QY 483 ATTCCGGAGTCTTGGGGCCGAGAAACCAAGTAGGAAACCCAGACATAGTAGGGTCTTC 542
 Db 481 ATTCCGGAGTCTTGGGGCCGAGAAACCAAGTAGGAAACCCAGACATAGTAGGGTCTTC 540
 QY 543 CTGTTGGGCAAGACACCCCCCAGAGGACCAACCCCTGACCTTTCCGACATGCTTAAAGTTT 602
 Db 541 CTGTTGGGCAAGACACCCCCCAGAGGACCAACCCCTGACCTTTCCGACATGCTTAAAGTTT 600
 QY 603 GAGATGACAAAGCACCAGGACATCCTCATGTGGAATATAGACACATCTTCAACCTG 662
 Db 601 GAGATGACAAAGCACCAGGACATCCTCATGTGGAATATAGACACATCTTCAACCTG 660
 QY 663 TCCTGAAGGAAGTCTGTTTCTTAGTGGTGGTGGAGCTTCCCTGTCAGACGAGAGTTT 722
 Db 661 TCCTGAAGGAAGTCTGTTTCTTAGTGGTGGTGGAGCTTCCCTGTCAGACGAGAGTTT 720
 QY 723 GTCTTCAAGGCGGATGATGACGTGTTTGTGAACACCCATCAGATCCTTAATTAAGT 782
 Db 721 GTCTTCAAGGCGGATGATGACGTGTTTGTGAACACCCATCAGATCCTTAATTAAGT 780
 QY 783 AGCTTATCCAGAGCAAGCAAGAGCTTCTTATAGTGGTGGTGGAGCTTCCACATGCTGG 842
 Db 781 AGCTTATCCAGAGCAAGCAAGAGCTTCTTATAGTGGTGGTGGAGCTTCCACATGCTGG 840
 QY 843 CCTCACCGGATAGAAACTGAAGTACTACATCCAGAACTTCTTACACCGGCTCTAC 902
 Db 841 CCTCACCGGATAGAAACTGAAGTACTACATCCAGAACTTCTTACACCGGCTCTAC 900
 QY 903 CCACCGTATCCCGGGGTGGTGGATTCCTGTACTCCGGCCCCCTTGCCTTGAGGCTGTAC 962
 Db 901 CCACCGTATCCCGGGGTGGTGGATTCCTGTACTCCGGCCCCCTTGCCTTGAGGCTGTAC 960
 QY 963 AGTCGACTAGCCGGGTCCATCTCTACCTTATGATGATGCTTTATACGGGAATGTCCTT 1022
 Db 961 AGTCGACTAGCCGGGTCCATCTCTACCTTATGATGATGCTTTATACGGGAATGTCCTT 1020
 QY 1023 CAGAACTGGGCTTGTTCAGAGAAAGCAAGGCTTCCAGGACATTTGATATTGAAGAG 1082
 Db 1021 CAGAACTGGGCTTGTTCAGAGAAAGCAAGGCTTCCAGGACATTTGATATTGAAGAG 1080
 QY 1083 AAAAATAGAAATATTTGTTCTATATAGACCTAATGTTAGTACATACGAAACCT 1142
 Db 1081 AAAAATAGAAATATTTGTTCTATATAGACCTAATGTTAGTACATACGAAACCT 1140
 QY 1143 CAAGAGATGATGATATCTGCTCTCAGTTCGAAAGTCTTAATTTAAATCTCTGA 1196
 Db 1141 CAAGAGATGATGATATCTGCTCTCAGTTCGAAAGTCTTAATTTAAATCTCTGA 1194

RESULT 7
 ABV75081
 ID ABV75081 standard; DNA; 1260 BP.
 XX ABV75081;
 XX
 DT 19-FEB-2003 (first entry)
 XX
 DE Murine beta3nT gene polynucleotide sequence.
 XX
 KW Beta3nT; transgenic; beta-1,3-N-acetylglucosaminyltransferase;
 XX disease system; gene; mouse; ds.
 OS Mus sp.

XX Key Location/Qualifiers
 FT CDS 30..1223
 FT /*tag= a
 FT /product= "beta3GnT"
 PN WC202279413-A2.
 XX 10-OCT-2002.
 XX 29-MAR-2002; 2002WO-US09645.
 XX 29-MAR-2001; 2001US-280706P.
 PR 28-MAR-2002; 2002US-0280706.
 XX (DELT-) DELTAGEN INC.
 XX Leviten MW, Phillips R;
 XX WPI: 2003-067437/06.
 DR P-PSDB; AB82648.
 XX New transgenic mouse comprising a disruption in a beta3GnT gene, as in
 PT vivo model to study various disease states or conditions in which
 PT beta3GnT may be implicated or involved, such as abnormal cell growth,
 PT cancer and metastasis -
 XX Disclosure: Fig 1; 55pp; English.
 XX The invention relates to a transgenic mouse comprising a disruption in a
 CC beta3GnT (beta-1,3-N-acetylglucosaminyltransferase) gene, where there is
 CC no native expression of beta3GnT gene. The transgenic mice may be used
 CC as in vivo model to study various disease states or conditions in which
 CC beta3GnT may be implicated or involved, such as abnormal cell growth,
 CC cancer and metastasis, and to evaluate various treatments or to identify
 CC agents for treating disease states or conditions, such as anxiety or
 CC depression. Animal-based disease systems may be used to identify
 CC compounds capable of ameliorating disease symptoms, as test substrates
 CC for the identification of drugs, pharmaceuticals, therapies and
 CC interventions that may be effective in treating a disease or other
 CC phenotypic characteristic of the animal. The present sequence represents
 CC a mouse beta3GnT polynucleotide sequence.
 XX
 SQ Sequence 1260 BP: 370 A; 282 C; 307 G; 301 T; 0 other;
 Query Match 99.7%; Score 1192.8; DB 25; Length 1260;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1194; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AGAAGAGTGTGGGCTCGAAGAGTCAGTTCCTGGGCATCCTGATGATGGCAATGTCT 60
 DB 28 AGAAGAGTGTGGGCTCGAAGAGTCAGTTCCTGGGCATCCTGATGATGGCAATGTCT 87
 QY 61 TCATTATTATTGATTGTGGAGTCTCCAAAACAGTAGCCACACAAAATGGAAAGGAG 120
 DB 88 TCATTATTATTGATTGTGGAGTCTCCAAAACAGTAGCCACACAAAATGGAAAGGAG 147
 QY 121 GAGTAATAATCCGAAAGAGAGTTCGGAAGCCACCCAGCTCCCGGGGCATCTCTGA 180
 DB 148 GAGTAATAATCCGAAAGAGAGTTCGGAAGCCACCCAGCTCCCGGGGCATCTCTGA 207
 QY 181 ACAGGGAACAGGAGAGTTCGGAAGCCACCCAGCTCCCGGGGCATCTCTGA 240
 DB 208 ACAGGGAACAGGAGAGTTCGGAAGCCACCCAGCTCCCGGGGCATCTCTGA 267
 QY 241 AGACAGGGAGCTAGCCATCTCCAAACACAGTCACTGAGCTATTGTGAACAGACT 300
 DB 268 AGACAGGGAGCTAGCCATCTCCAAACACAGTCACTGAGCTATTGTGAACAGACT 327
 QY 301 CGACGGTATGACAGCTGTGACAGATTTTAATTAATCTGCCGACAGATTTAAGACTTTC 360
 DB 328 CGACGGTATGACAGCTGTGACAGATTTTAATTAATCTGCCGACAGATTTAAGACTTTC 387

QY 361 TCTTGTATTGAGATGCCGGAATTAATCTCGTCTTATAGATCAACCCGAAGAAATGTGCAA 420
 DB 388 TCTTGTATTGAGATGCCGGAATTAATCTCGTCTTATAGATCAACCCGAAGAAATGTGCAA 447
 QY 421 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAG 480
 DB 448 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAG 507
 QY 481 CAATTGGGAGTCTTTGGGCGCAGAGAAACCAAGTAGGGAACAGACAGTACTGAGGGTCT 540
 DB 508 CAATTGGGAGTCTTTGGGCGCAGAGAAACCAAGTAGGGAACAGACAGTACTGAGGGTCT 567
 QY 541 TCTTGTGGCAAGACACCCCGAGAGACACACCCCTGACCTTTTGGACAIGCTTAAGT 600
 DB 568 TCTTGTGGCAAGACACCCCGAGAGACACACCCCTGACCTTTTGGACAIGCTTAAGT 627
 QY 601 TTGAGAGTCAAGAGCAGCAGGACATCTCTATGTGGAACATATAGACACATCTTCAACC 660
 DB 628 TTGAGAGTCAAGAGCAGCAGGACATCTCTATGTGGAACATATAGACACATCTTCAACC 687
 QY 661 TGTCCCTGAAGAGTGTCTTTTCTTAGTGGGTGAGCACTTCTCTGTCACAGCAGAGT 720
 DB 688 TGTCCCTGAAGAGTGTCTTTTCTTAGTGGGTGAGCACTTCTCTGTCACAGCAGAGT 747
 QY 721 TGTCTTCAAGGCGGATGATGACGTGTTTGTGAACACCCATCAGATCTTAACTACTGA 780
 DB 748 TGTCTTCAAGGCGGATGATGACGTGTTTGTGAACACCCATCAGATCTTAACTACTGA 807
 QY 781 ATAGCTTATCCAGAGCAAGCCAAAGACTTGTTCATAGGTGACGTGATCCACAATGCTG 840
 DB 808 ATAGCTTATCCAGAGCAAGCCAAAGACTTGTTCATAGGTGACGTGATCCACAATGCTG 867
 QY 841 GGCTCACCAGGATGAAGAACTGAAGTACTACATCCAGAGTCTTCTACACCCGGCTCT 900
 DB 868 GGCTCACCAGGATGAAGAACTGAAGTACTACATCCAGAGTCTTCTACACCCGGCTCT 927
 QY 901 ACCACCGTATGCGGGGTGGTGGATCTCTACTTCGGGCCCCCTTGCCTTGAGGGTGT 960
 DB 928 ACCACCGTATGCGGGGTGGTGGATCTCTACTTCGGGCCCCCTTGCCTTGAGGGTGT 987
 QY 961 ACAGTCCGACTAGCCGGGTCCATCTCTACCTTATTGATGATGTTTATACGGGAATGTGCC 1020
 DB 988 ACAGTCCGACTAGCCGGGTCCATCTCTACCTTATTGATGATGTTTATACGGGAATGTGCC 1047
 QY 1021 TTCAGAACTGGGCCCTTGTTCAGAGAGACACAAAAGGCTTCAGGACATTTGATATTGAAG 1080
 DB 1048 TTCAGAACTGGGCCCTTGTTCAGAGAGACACAAAAGGCTTCAGGACATTTGATATTGAAG 1107
 QY 1081 AGAAAAATAGAANAATATTGTTCTATATAGACCTTAATGTTAGTACATAGCAGAAAAC 1140
 DB 1108 AGAAAAATAGAANAATATTGTTCTATATAGACCTTAATGTTAGTACATAGCAGAAAAC 1167
 QY 1141 CTCAGAGATGATTGATATCTGTGCTCAGTTGCAAAAGTCTTAATTTAAATGCTGA 1196
 DB 1168 CTCAGAGATGATTGATATCTGTGCTCAGTTGCAAAAGTCTTAATTTAAATGCTGA 1223
 RESULT 8
 ID AAA12716 standard; cDNA: 1707 BP.
 XX AC AAA12716;
 XX AC AAA12716;
 XX 25-JUL-2000 (first entry)
 XX cDNA encoding human LIG46 polypeptide.
 DE Leptin; LIG46; body weight; leptin inducible gene; obesity; cachexia;
 KW LIG56; Tgtp; LRG-47; RC10-II; Stral3; ss.
 XX Homo sapiens.
 XX OS
 XX Key Location/Qualifiers

```

FT CDS 245..1439
FT /*tag= a
FT /product= "LIG46"
XX WO200015826-A2.
XX 23-MAR-2000.
XX 10-SEP-1999; 99WO-US20722.
XX 10-SEP-1998; 98US-0150857.
XX 29-OCT-1998; 98US-0106378.
XX 19-NOV-1998; 98US-0195896.
XX 15-APR-1999; 99US-0292228.
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX White D, Zhou J, Tartaglia LA;
XX
XX WPI; 2000-271461/23.
XX P-PSDB; AAY84683.
XX
XX Method for determining compounds which modulate body weight and can be
XX used to treat e.g. obesity comprises measuring the activity of leptin;
XX Inducible genes
XX
XX Disclosure; Fig 7; 123pp; English.
XX
XX The present sequence encodes a leptin induced LIG46 polypeptide. The
XX specification describes a method for determining whether a compound can
XX be used to modulate body weight by measuring the activity of leptin
XX inducible genes, such as LIG46. The method can be used to specifically
XX detect LIG46 and LIG56 nucleic acid molecules relative to other nucleic
XX acid molecules encoding galactosyltransferases or GTP-binding proteins.
XX They can also be used in diagnostic assays to identify the presence or
XX absence of a genetic lesion or mutation characterized by aberrant
XX modification or mis-regulation of the genes or aberrant post-
XX translational modification of the proteins. LIG46 and LIG56 proteins
XX and nucleic acid molecules can be used to treat obesity or cachexia.
XX LIG46 and LIG56 antagonists are used to treat obesity and LIG46 and
XX LIG56 agonists are used to treat low body weight. The leptin inducible
XX genes i.e. LIG46, LIG56, Tgtp, LRG-47, RC10-II and Stral3 can be used
XX to determine whether a compound modulates body weight and can then be
XX used to treat obesity or cachexia or low body weight.
XX
XX Sequence 1707 BP: 452 A; 403 C; 435 G; 400 T; 17 other;
XX
XX Query Match 75.8%; Score 906.4; DB 21; Length 1707;
XX Best Local Similarity 84.9%; Pred. No. 5.6e-270;
XX Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;
XX
XX 1 AGATGAGTGTGGGCGCTCGAAGAGTCAGTTCCTGGCGATCCCTGATGTCGAAATGTCT 60
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 244 AAATGAGTGTGGAGCTCGAAGATAAAGTTGTTGGGTATCCTGATGTCGAAATGTCT 303
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 61 TCATTTATTTGATTTGTGAAGTCTCCAAAACAGCTAGCAGACAAAATGGAAGGAG 120
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 304 TCATTTATTTATGAGTCTCCAAAACAGCTAGCAGACAAAATGGAAGGAG 363
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 121 GAGTAAATATCCGAAAGAGAGTTCCTGGAAGCCAGCCAGATCCCGGGGCATCTGGA 180
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 364 AAGTAAATATCCCAAGAGAGTTCCTGGAAGATATCTACCCCTCCCGAGGCATCTGGA 423
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 181 ACAGGGAACAGAGAGTGTGAAGTGTGATACATCTCTTCAACAGGTTGCCCAATC 240
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 424 ACCGAGACAGAGAGTGTGAAGTGTGATACATCTCTTCAACAGGTTGCCCAATC 483
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 241 AGACAGGGGAGCTAGCCATCTCCAAACAGATCACTGAGCTATTGTGAACAGACT 300
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 484 AGACAGGGGAGCGGCGAGGCTCTCCAAATATGAAGCATCTGAACACTGCGAAGCTGACC 543
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 301 CGACGGTATACAGCTGTGACAGATTTTAAATATCTGCGGACAGAGATTTAAAGACTTTC 360
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

544 TGAGGTCACGTCGGTGGTTACGGGTTTAAACAACCTTGCAGACAGATTTAAAGACTTTC 603
361 TCTTGTATTGAGATCGCGGAATTAACGCTGCTTATAGATCAACCGAAGAAATGTGCRA 420
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
604 TGCTGTATTGAGATCGCGCAATTAATTCACCTGCTTATAGATCAGCCGGATGATGTCGA 663
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
421 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTTCCACATTTTGCAGAGCAAG 480
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
664 AGAAGCTTCTTCTGCTGGCGATTAACTCCCTCACTCCACATTTTGCAGAGCAAG 723
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
481 CAATTCGGGAGTCTTGGGGCGAGAAACCAACAGTAGGAAACACAGATAGTACAGGCTCT 540
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
724 CAATCCGGGAATCTCTGGGGCCCAAGAAAGCAACCGAGGAAACCAAGGCTGTCGAGTCT 783
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
541 TCCTGTTGGGCAAGACACACCCCGAGAGCAACACCTGACCTTTCGGACATGCTTAAGT 600
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
784 TCCTGCTGGCCAGACACACCCCGAGAGCAACACACCCCGACCTTTCAGATATGCTGAAT 843
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
601 TTGAGAGTGACAGACACACAGACATCTCATGTGGAACTATAGAGACACATTTCTCAACC 660
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
844 TTGAGAGTGAGAGCAACCAAGACATCTTATGTGGAACACACACCCCGACCTTCTTCAACT 903
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
661 TCTCCCTCAAGGAAGTCTGTTTCTTAGTGGGTGAGCACTTCTGTCAGAGACGACAGT 720
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
904 TGTCTCTGAAGGAAGTGTCTTCTCAGGTGGGTGAAGTACTTCTGCCCAGACACATGAGT 963
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
721 TTGCTTCAAGGGCGATGATGACGTGTTTGTGAACACCCCATCACATCTTAAATCTTGA 780
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
964 TTGTTTCAAGGGCGATGACCATGTTTGTGAACACCCCATCACATCTTAAATCTTGA 1023
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
781 ATAGCTTATCAAGAGCAACCCCAAGACTTGTTCATAGGTGAGCTGATCCCAATGCTGT 840
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1024 ATAGTTTATCAAGACCCCAAGACTTCTTTCATAGGTGATGTCATCCCAATGCTGT 1083
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
841 GGCCTCACCGGGATGAAGAACTGAACTACTACATCCCAAGAGTCTTCTACACGGGCTGT 900
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1084 GACCTCATCGGATGAAGAGCTGAAGTACTACATCCCAAGAGTGTGTTACTCTGGCTCT 1143
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
901 ACCCACTATCGCGGGGTGGTGGATTCCTGTACTCCGGCCCTTGCCTTGAGGCTGT 960
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1144 ACCCACTATCGAGGGGAGGGGGTCTCTACTCCGGCCACTGGCCCTGAGGCTGT 1203
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
961 ACAGTGCACATGAGCGGGTCTCTACCTTATGATGATGTTTATACGGGAATGTC 1020
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1204 ACCATATCACTGACACAGGTCTCTACCCCATGATGAGCTTACTGGATGTGCC 1263
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1021 TFCAGAACTGGGCTTGTCCAGAGAGCAACAAAGGCTTCAGACATTTGATTTGAAG 1080
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1264 TTCAGAACTGGGCTTGTCCAGAGAAACAAAGGCTTCAGACATTTGATTCGAGG 1323
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1081 AGAAAAATAAGAAAAATATTTTCTTATAGACCTAATGTTAGTACATAGCAGAAAC 1140
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1324 AGAAAAATAAGAAAAATATTAACATCTGCTATGATGATGATGTTAGTACATAGTAAAC 1383
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1141 CTCAGAGATGATTCATGATCTGCTCAGTTTGCAGGTCGAAAGTCTTAATTTAAATGCTGA 1196
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1384 CTCAGAGATGATTCATGATTTGGTCTCAGTGCAGAGTGTCTATTAAATGCTAA 1439
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 9
AAI67869
ID AAI67869 standard; cdna; 1707 BP.
XX
AC AAI67869;
XX
XX 13-MAR-2002 (first entry)
XX
XX Human LIG46 polypeptide encoding cdna.
XX
XX Leptin; LIG46; LIG56; leptin induced gene; Tgtp; LRG-47; RC10-II; Stral3;
XX anorectic; anabolic; antisense therapy; human; ss.
XX
XX Homo sapiens.
OS

```



```

FT      /*tag= a
FT      /product= "Human LIG46 protein"
FT      246...341
FT      *tag= b
FT      342...1436
FT      /tag= c
FT      /product= "Human mature LIG46 protein"
XX      WC200274905-A2.
XX      26-SEP-2002.
XX      20-NOV-2001; 2001WO-US43345.
XX      21-NOV-2000; 2000US-0717778.
XX      (MILL-) MILLENNIUM PHARM INC.
XX      White DW, Zhou J, Tartaglia LA, Stricker-krongrad A, Clausen H;
XX      WPI; 2002-759886/82.
XX      P-PSDB; AAE29093.
XX      New isolated nucleic acid molecules, designated as leptin-induced genes
XX      46 (LIG46), useful for treating a subject having a disorder
XX      characterized by undesirable level of LIG46 expression or activity.
XX      such as low body weight
XX      Claim 15; Fig 5; 90pp; English.
XX      The invention relates to LIG46, a gene whose expression is induced by
XX      leptin. LIG46 DNA and protein are useful in treating a subject having
XX      a disorder characterized by undesirable level of LIG46 expression or
XX      activity, such as low body weight. They are also useful in a screening
XX      assay, chromosomal mapping, tissue typing and forensic biology. The
XX      probes based on the LIG46 nucleotide sequence are useful for detecting
XX      transcripts or genomic sequences encoding the same or related proteins.
XX      The LIG46 polypeptides are useful as immunogens for raising anti-LIG46
XX      antibodies. The host cells are useful for producing non-human transgenic
XX      animals. LIG46 DNA is used in gene therapy. The present sequence is
XX      human LIG46 cDNA. LIG46 gene is located at chromosome 2.
XX      Sequence 1707 BP; 452 A; 403 C; 435 G; 400 T; 17 other;
XX      Query Match 75.8%; Score 906.4; DB 24; Length 1707;
XX      Best Local Similarity 84.9%; Pred. No. 5.6e-276;
XX      Matches 1045; Conservative 0; Mismatches 181; Indels 9; Gaps 0;
OY      1 AGATGAGTGTGGGGCGTGAAGAGTCAAGTTGCTGGGCATCCTGATGATGCAAAATGTCT 60
Db      244 AATGAGTGTGGAGCTGCAAGAAATAAAGTTGTTGGGTATCCTGATGATGCAAAATGTCT 303
OY      61 ICATTTATTATTGTTGGAGGTCCTCAAAACAGTAGCCAAAGAAATAATGGAAGGGAG 120
Db      304 TCATTTATTATTATGGAAGTCTCCAAAGCAGTAGCCAAAGAAATAATGGAAGGGG 363
OY      121 GASTAATAATCCGAAAGAGAGTCTTGAAGCCACCCAGCAGTCTCCCGGCACTATGGA 180
Db      364 AAGTAATAATACCAAGAGAGAGTCTTGAAGATACTACCCCTCCCGAGCAGTATGGA 423
OY      181 ACAGGGAACAGAGAGAGTGAACAGGTGGTACAAATCCCATCTTGAACAGGTGGCCAAATC 240
Db      424 ACCGAGAGCAAGAGAGTGAAGTGAAGTGGTGAAGTGGTGAAGTGGTGAAGTGGTGAAGT 483
OY      241 AGACAGGGGAGCTAGCCATCTCCAAACACAAAGTCAGTCTGATGATGTTGTTGAACAGACT 300
Db      484 AGACGGGGAGGGCGGCGAGGCTCTCCAAATATAGGCATCTGACTACTGCGAAGCTGACC 543
OY      301 CGAGGTGATGACAGCTGTGACAGATTTTAAATATCTCCCGGACAGATTTAAAGACITTC 360
Db      544 TGAGGGTCAAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 603
OY      361 TCTTGTATTGTAGATCCCGAATTACTCGCTCTTAAGATCAACCCGAGAAATGTGCAA 420

```

```

Db      604 TCCTGTAITTTGAGATGCCGCAATATTTCACATGCTATATAGATCAGCCGGATAGTGTCAA 663
OY      421 AGAAGCCCTTCTTACTATTGGGATAAAGTCCCTCATTTCCACATTTTGGCCAGAGGAAG 480
Db      664 AGAAACCTTCTTGTGTTGCTGGCGATTAAAGTCCCTCACTCCACATTTTGGCAGAGGAAG 723
OY      481 CAATTCGGGAGTCTTGGGGCGGAGAAACCAAGTAGGGAACACACAGTAGTAGGGGTCT 540
Db      724 CAATTCGGGAATCTTGGGGCCAGAAAGACACAGGGAACCAAGGTTGGTGGAGTCT 783
OY      541 TCTGTGGGCAAGACACCCCGAGAGAACACCCCTGACCTTTCGGACATGCTTAAGT 600
Db      784 TCCGTGTCGGCAGACACCCCGAGAGAACACCCCGAGACCTTTCAGATATGCTGAAT 843
OY      601 TTGAGAGTGACAAAGCAGGAGACATCTCATGTGGAACATATAGAGACACATCTTCAACC 660
Db      844 TTGAGAGTGAGAAAGCAGGACCAAGACATCTTATGTGGAACATACAGAGACATTTCTTCACT 903
OY      661 TGTCCCTGAAGGAAGTGTCTTGTAGTGGGTGAGCAGTTCCTGTCCAGACGACAGAGI 720
Db      904 TGTCTCTGAAGGAAGTGTCTTGTAGTGGGTGAGTAACTTCTCTGCTGCCAGACACTGAGT 963
OY      721 TTGTCTTCAAGGCGCATGATGACGTGTTTGTGAACACCCATCAGATCTCTTAATTAAGTGA 780
Db      964 TTGTCTTCAAGGCGCATGACATGTTTGTGAACACCCATCAGATCTCTTAATTAAGTGA 1023
OY      781 ATAGCTTATCCAAAGAGCAAGCAAGACTTGTTCATAGTGAAGTGTATCCACAAATGCTG 840
Db      1024 ATAGTTTATCCAAAGACCAAGCAAGACTTCTCTCATAGTGTATGATCCCAATGCTG 1083
OY      841 GGCCTCACCGGATGAAGAACTGAAGTACTACTATCCAGAGAGTCTTCTACACCGGCTCT 900
Db      1084 GACCTCATCGGATGAAGAGCTGAAGTACTACTATCCAGAGAGTGTCTTACTCTGGCCTCT 1143
OY      901 ACCACCGTATCCCGGGGTGGTGGATTCCTGTACTCCGGCCCTTCCCTTGAGGCTGT 960
Db      1144 ACCACCGTATCCCGGGGTGGTGGATTCCTGTACTCCGGCCCTTCCCTTGAGGCTGT 1203
OY      961 ACAGTCCGACTAGCCGGGTCCATCTCACTACCTATGATGATGTTTATACGGGAATGTGCC 1020
Db      1204 ACCATATCACTGACCAAGTCCATCTCTACCCCAITGTAGTACGTTATATGATGTGCC 1263
OY      1021 TTCAGAAACTGGGCTTGTTCAGAGAGAGCAAGAGGCTTCAGGACATTTGATATGAAG 1080
Db      1264 TTCAGAAACTGGGCTTGTTCAGAGAGAGCAAGAGGCTTCAGGACATTTGATATGAGG 1323
OY      1081 AGAAATAAAGAAATAATTTCTTCTATATAGACCTAATGTTAGTACATAGCAAGAAAC 1140
Db      1324 AGAAATAAAGAAATAATTTCTTCTATATAGACCTAATGTTAGTACATAGCAAGAAAC 1383
OY      1141 CTCAGAGATGATGATATGATATCTGCTCAGTTCAGAAAGTCCCTAAATTTAAATGCTGA 1196
Db      1384 CTCAGAGATGATGATATTTGCTCAGTTCAGAGTGCATGCTCAATTTAAATGCTAA 1439

```

RESULT 11

ABV75082

ID ABV75082 standard; DNA; 1831 BP.

XX AC ABV75082;

XX DT 19-FEB-2003 (first entry)

XX XX Human beta3nT gene polynucleotide sequence.

XX XX Beta3nT; transgenic; beta-1,3-N-acetylglucosaminyltransferase;

XX XX disease system; gene; human; ds.

XX OS Homo sapiens.

XX XX Key Location/Qualifiers

XX FT CDS 235..1429

```
FT      /*tag- a
PN      /product= "beta3GnT"
XX      W0200279413-A2.
XX      10-OCT-2002.
PD      25-MAR-2002; 2002MO-US09645.
XX      29-MAR-2001; 2001US-280706P.
XX      28-MAR-2002; 2002US-0280706.
XX      (DELTA-) DELTAGEN INC.
XX      Leviten MW, Phillips R;
PI      WPI; 2003-067437/06.
XX      P-PSDB; ABB82649.
DR      New transgenic mouse comprising a disruption in a beta3GnT gene, as in
XX      vivo model to study various disease states or conditions in which
PT      beta3GnT may be implicated or involved, such as abnormal cell growth,
PT      cancer and metastasis.
XX      Disclosure; Fig 3; 55pp; English.
XX      The invention relates to a transgenic mouse comprising a disruption in a
CC      beta3GnT (beta-1,3-N-acetylglucosaminyltransferase) gene, where there is
CC      no native expression of beta3GnT gene. The transgenic mice may be used
CC      as in vivo model to study various disease states or conditions in which
CC      beta3GnT may be implicated or involved, such as abnormal cell growth,
CC      cancer and metastasis, and to evaluate various treatments or to identify
CC      agents for treating disease states or conditions, such as anxiety or
CC      depression. Animal-based disease systems may be used to identify
CC      compounds capable of ameliorating disease symptoms, as test substrates
CC      for the identification of drugs, pharmaceuticals, therapies and
CC      interventions that may be effective in treating a disease or other
CC      phenotypic characteristic of the animal. The present sequence represents
CC      a human beta3GnT polynucleotide sequence.
XX      Sequence 1831 BP; 497 A; 423 C; 443 G; 468 T; 0 other;
SQ      Query Match      75.8%; Score 906.4; DB 25; Length 1831;
      Best Local Similarity 84.9%; Pred. No. 5.8e-270;
      Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;
QY      1 AGATGAGTGGGGCGTGAAGAGTCAAGTGTGCGCATCCCTGATGAGCGAAATGTCT 60
DB      11111111 11 11111111 11111111 11111111 11111111 11111111 11111111
QY      61 TCATTATTTTGAATGTGGAGTCTCCAAACACAGTAGCCCAAGACAAAATGGAAAGGGAG 120
DB      11111111 11 11111111 11111111 11111111 11111111 11111111 11111111
QY      294 TCATTATTTTATATGAAGTCTCCAAAGAGCAGTAGCCCAAGAAAATGGAAAGGGG 353
DB      11111111 11 11111111 11111111 11111111 11111111 11111111 11111111
QY      121 GAGTAATATCCCAAGAGAGAGTCTGGAGGCCACCCAGCAGCTCCCGGGGCAATCTGA 180
DB      11111111 11 11111111 11111111 11111111 11111111 11111111 11111111
QY      354 AAGTAATAATACCCAAAGAGAGTCTGGAGAGATATCTACCCCTCCCGAGGCACTCTGA 413
DB      11111111 11 11111111 11111111 11111111 11111111 11111111 11111111
QY      181 ACAGGGAACAGGAGAGCTGAACAGGTGGTACAAATCCATCTTGAACAGGGTGGCCAAATC 240
DB      11111111 11 11111111 11111111 11111111 11111111 11111111 11111111
QY      414 ACCGAGACCAAGAGAGTGAACCGGAGTACACACCCATCTCTGAGCATGCTGACCAACC 473
DB      11111111 11 11111111 11111111 11111111 11111111 11111111 11111111
QY      241 AGACAGGGGAGCTAGCCACATCTCCAAACACAGTCACTGAGCTATTGTGAACACAGACT 300
DB      11111111 11 11111111 11111111 11111111 11111111 11111111 11111111
QY      474 AGACGGGGGAGCGGGAGGCTCTCCAAATATAGCCCATCTGAACCTACTGCGAACCCTGACC 533
DB      11111111 11 11111111 11111111 11111111 11111111 11111111 11111111
QY      301 CGACGGTCATCAGAGCTGTGACAGATTTTAATACTCTGCCGAGAGATTTAAGAGCTTC 360
DB      11111111 11 11111111 11111111 11111111 11111111 11111111 11111111
QY      534 TGAGGGTCACGTCGGTGTACGGGTTTACAACTTTCGCGGACAGATTTAAGAGCTTC 593
DB      11111111 11 11111111 11111111 11111111 11111111 11111111 11111111
QY      361 TCATTATTTAGATCGCGGAATTTACTCGCTGCTTATAGATCAACCGAAGAAATGTGCAA 420
DB      11111111 11 11111111 11111111 11111111 11111111 11111111 11111111
QY      594 TGCTGTATTTGAGATCGCGCAATTTATTCAGTCTTATAGATAGCCCGGATAAGTGTGCA 653
DB      11111111 11 11111111 11111111 11111111 11111111 11111111 11111111
```

```
QY      421 AGAAGCCCTCTTACTATTGGCGATAAAGTCCCTCATTTCCACATTTTGGCAGAGCAAG 480
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
DB      554 AGAAGCCCTCTTCTTGTGCGGATTAAGTCCCTCACTCCACATTTTGGCAGAGCAAG 713
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
QY      481 CAATTCCGGGAGTCTTGGGGCCGAGAAACCAACGTAAGGAACACAGACAGTAGTAGGCTCT 540
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
DB      714 CAATCCGGGAATCCTGGGGCCCAAGAAAGCAACGACGGAACCAACGGTGGTGGAGTCT 773
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
QY      541 TCCTGTTGGCAGACACACCCCGCAGAGCAACACCCCTGACCTTTCCGACATGCTCTAACT 600
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
DB      774 TCCTGCTGGGCGCAGACACCCCGCAGAGCAACACCCCGCAGCTTTCAGATATGCTGAAT 833
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
QY      601 TTGAGAGTGACAAAGCACACAGGACATCCTCATGTGGAACATATAGAGACACATCTTCAACC 660
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
DB      834 TTGAGAGTGAGAAAGCACCAAGACATCTTATGTGGAACACTACAGAGACACTTCTTCAACT 893
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
QY      661 TGTCCCTGAAGGAGTGTCTTTCTAGTGGGTGAGCACTTCTCCTCCAGAGCGAGAGT 720
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
DB      894 TGTCTCTGAAGGAGTGTCTTTCTCAGGTGGGTAAAGTACTTCCCTGCCGAGACACTGAGT 953
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
QY      721 TTGTCTTCAAGGGCGATGATGAGTGTGTTGTGAACACCCATCACATCTTAACTTGA 780
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
DB      954 TTGTTTTCAAGGGCGATGAGTGTGTTTGTGAACACCCATCACATCTCTGAATTTCTGA 1013
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
QY      781 ATAGCTTATCCAGAGCAACCAAGCACTTGTTCATAGTGTGAGCTGATCCACATGCTG 840
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
DB      1014 ATAGCTTATCCAGACCAACCAAGCACTTTCATAGGTGAAGTGTGATCCACATGCTG 1073
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
QY      841 GGCCTCACGGGATAAAGAACTGAAGTACTACATCCAGAAAGTCTTCTACACCGGGCTCT 900
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
DB      1074 GACCTCATCGGATAGAAGCTGAGTACTACATCCAGAAAGTGTGTTACTCTGGCTCT 1133
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
QY      901 ACCCACGATATCCGGGGGTGGTGGATTCCTGATCTCCGGCCCCCTTGGCTTGAAGCTCT 960
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
DB      1134 ACCCACCTATGACGGGGGAGGGGGTTCCTCTACTCCGGCCACCTGGCCCTGAGCTCT 1193
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
QY      961 ACATGGCATGAGCGGGTCCATCTCTACCCCTATTGATGATGTTTATACGGGAATGCTGC 1020
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
DB      1194 ACCATATCATGACCGAGTCCATCTCTACCCCATGATGACGTTTATCTGGAAATGGCC 1253
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
QY      1021 TTCAGAAACTGGGCTTGTTCAGAGAAAGCAAAAGGCTTCAGGACATTTGATATTGAAG 1080
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
DB      1254 TTCAGAAACTCGGCTCTGTTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATCGAGG 1313
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
QY      1081 AGAAATTAAGAAAAATTTGTTCTATATAGCACTTAAGTGTAGTACATAGCAGAAAC 1140
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
DB      1314 AGAAAAACAAAAATAACATCTGCTCTATGTAGATCTGATGTTAGTACATAGTAGAAAC 1373
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
QY      1141 CTCAGAGATGATGATGATCTGCTCAGTTTGCAGTTTGCAGAAAGTCCATAATTTAAATGCTGA 1196
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
DB      1374 CTCAGAGATGATGATGATTTGTTCTCAGTTTGCAGAGTGTCTCATTTAAATGCTAA 1429
XX      1111 111111 11111111 11111111 11111111 11111111 11111111 11111111
RESULT 12
AAF29255
ID      AAF29255 standard; cDNA; 1912 BP.
XX      AAF29255;
XX      AC
XX      17-APR-2001 (first entry)
XX      Human beta 1,3-N-acetylglucosamine transferase G3 cDNA sequence.
DE      Sugar chain synthesis agent; beta 1,3-N-acetylglucosamine transferase;
KW      inflammation; cancer; metastasis; human; ss.
XX      Homo sapiens.
OS      Homo sapiens.
XX      WO200100848-A1.
XX      PD
XX      04-JAN-2001.
XX
```

29-JUN-2000; 2000WO-JP04304.
29-JUN-1999; 99JP-0283437.
16-MAR-2000; 2000JP-0074757.
(KYOW) KYOWA HAKKO KOGYO KK.
Sasaki K, Shiraishi N, Natsume A, Yamada Y, Nakagawa S, Sekine S;
WPI: 2001-102895/11.
P-PsDB: AAB49748.
XX New polypeptide having betal,3-N-acetylglucosamine transferase activity
PT for diagnosis of inflammation, cancer and cancer metastasis, -
PT development of remedies, and for producing glycoconjugates
XX
XX Example 2: Page 159-164; 195pp; Japanese.
XX This invention relates to a sugar chain synthesizing agent that contains
CC a polypeptide as the active ingredient, where the polypeptide has beta
CC 1,3-N-acetylglucosamine transferase activity. The polypeptide has beta
CC sequences encoding them and antibodies directed against the proteins are
CC useful in the diagnosis of inflammation, cancer and its metastasis,
CC development of remedies, and for producing sugar chains and
CC glycoconjugates. Sequences AAB49748 - AAB49751 represent polypeptides of
CC the invention, having beta 1,3-N-acetylglucosamine transferase activity.
CC and polynucleotides AAF29255 - AAF29258 represent cDNA encoding them. PCR
CC primers used in the isolation and characterisation of the cDNA sequences
CC are represented by sequences AAF29259 - AAF29290.
XX
SQ Sequence 1912 BP; 518 A; 430 C; 457 G; 507 T; 0 other;
Query Match 75.8%; Score 906.4; DB 22; Length 1912;
Best Local Similarity 84.9%; Pred. No. 66-270;
Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 3;
QY 1 AGATGAGTGTGGGGCGTCAAGAGTCAAGTGTCTGGGCACTCTGATGCGCAAAATGTCT 60
DB 232 AATGAGTGTGGAGCTGCAAGAAATAAGTGTGGGTATCTCTGATGCGCAAAATGTCT 291
QY 61 TCATTTATTTGATTTGGAAGTCTCCAAACACATAGCCAGCAAAATGGAAGGGAG 120
DB 292 TCATTTATTTTATTAAGTGTCTCCAAAGAGTGTAGCCAGCAAAATGGAAGGGG 35;
QY 121 GAGTAAATATCCGGAACAGAGTGTCTGGAGCCACCCAGCACTCCCGGGCATCTGGA 180
DB 352 AAGTAAATATACCCAAAGAGAGTGTCTGGAAGATATCTACCCCTCCCGAGGCATCTGGA 411
QY 181 ACAGGGAACAGAGAGAGTCTCAACAGCAAGTCACTGAGTATGTCAACAGACT 300
DB 412 ACCGAGAGCAAGAGAGTGTGAACCGGAGTACAAACCCCATCTCTGAGCATGCTGACCAAC 471
QY 241 ACAGAGGGAGTACCCAGTCTCCAAACAGTCACTGAGTATGTCAACAGACT 300
DB 472 ACAGGGGGAGGGGGAGGCTCTCCAAATATAAGCCATCTGAACTCTGCGAAGCTGACC 531
QY 301 CCAGCGTCATGACAGCTGTGACAGATTTTAAATATCTGCGGACAGATTTAAAGACITTC 360
DB 532 TCAGGGTCACGTCGGTGTACGGGTTTAAACACTTCCCGGACAGATTTAAAGACTTC 591
QY 361 TCTTGTATTTAGATGCGGGAATTTACTGCTGCTTTATAGATCAACCGAAGAAATGTGCAA 420
DB 592 TGCTGTATTTGAGATGCGGCAATTTATTCACCTGCTTATAGATCAGCGGATTAAGTGTGCAA 651
QY 421 ACAAGCCCTTCTTACTATTGGGATAAGTCCCTCATTCACATTTTCCGAGAGGCAAG 480
DB 652 AGAAACCTTCTTGTGTCTGGGATTAAGTCCCTCATTCACATTTTCCGAGAGGCAAG 711
QY 481 CAATTCGGGAGTCTTTGGGGCCGAGAAACACAGTAGGGAACACAGACAGTAGTAGGGTCT 540
DB 712 CAATCCGGGAATCTTGGGGCCGAGAAACACAGTAGGGAACACAGCGGTGGTGGAGTCT 771
QY 541 TCCTGTTGGGCAAGACACCCCGAGAGGACACACCCCTGACCTTTTCGACATGCTTAAGT 600

DB 772 TCCTGCTGGGCGACAGACACCCCGACAGGACAAACCCCGACCTTTCAGATATGCTGAAAT 831
QY 601 TTGAGAGTGACAAGCACCAGGACATCTCATCTGGAACACTATAGAGACACACTTCTTCAACC 660
DB 832 TTGAGAGTGAGAAGCACCAGGACATCTTATGTGGAACACTACAGAGACACTTCTTCAACT 891
QY 661 TGTCCTCAAGCAAGTGTCTTCTTAGGTGGGTGAGCACTTCTGTCCAGAGCGCAGAGT 720
DB 892 TGTCTCTCAAGCAAGTGTCTTCTCAGGTGGGTAGTACTTCTGCCACAGACACTGAGT 951
QY 721 TTGCTTCAAGGGCGATGATGACCTGTTGTGAACACCCATCACATCTTAAATTAATTGA 780
DB 952 TTGTTTCAAGGGCGATGACGATGTTTGTGAACACCCATCACATCTGAAATTAATTGA 1011
QY 781 ATAGCTTATCAAGCAGCAAGCCAAAGACTTGTTCATAGGTGAGCTGATCCCAANTGCTG 840
DB 1012 ATAGTTTATCAAGCAGCAAGCCAAAGACTTCTTCATAGGTGATGATCCCAANTGCTG 1071
QY 841 GGCCTCACCGGATTAAGAACTGAAGTACTACATCCAGAGTCTTCTACACCGGCTCT 900
DB 1072 GACCTCATCGGATTAAGAACTGAAGTACTACATCCCAAGAGTGTGTTACTCTGGCCTCT 1131
QY 901 ACCCACCTATGCGGGGGTGTGATTCCTGTACTCCGCCCTTGCCTTGAGGCTGT 960
DB 1132 ACCCACCTATGCGAGGGGGGGGTTCCTTCTACTCCGCCACCTGCGCCTGAGGCTGT 1191
QY 961 ACAGTGCAGTACGCGGGTCCATCTCTACCTATGTATGATGTTTATACGGGATGTGCC 1020
DB 1192 ACCATATCACTGACGAGGTCCTCTACCCATTTAGTACGTTTATATCTGGAATGTGCC 1251
QY 1021 TTCAGAAACTGGGCTTCTTCAGAGAGCAGCAAGGCTTCAGGACATTTGATATTGAAG 1080
DB 1252 TTCAGAAACTGGGCTTCTTCAGAGAAACAGCAAGGCTTCAGGACATTTGATATCAGG 1311
QY 1081 AGAAAAATTAAGAAAAATTTTCTCTATATAGACCTAAATTTAGTACATAGCAGAAAAAC 1140
DB 1312 AGAAAAACAAAAATAACATCTGCTCTATGTAGATCTGATGTTAGTACATAGTAAAAAC 1371
QY 1141 CTCAGAGATGATGATATCTGCTCAGTTGCAAGTCTCTAAATTTAAATGCTGA 1196
DB 1372 CTCAGAGATGATGATATTTGGTCTCAGTTGCAAGTCTCTAAATTTAAATGCTGA 1427
RESULT 13
AAV6367
ID AAV6367 standard; cDNA: 2745 BP.
XX AAV6367;
AC AC
XX DT 18-JAN-1999 (first entry)
XX DE human cardiac and pancreatic protein (CAPP) gene.
XX KW Cardiac and pancreatic protein; CAPP; muscle-derived growth factor;
XX KW human; pancreatitis; myocardial infarction; cardiomyopathy; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT CDS 233..1426
FT sig_peptide /tag= a 233..328
FT mat_peptide /tag= b 329..1423
FT /tag= c
XX PN W09844112-A1.
XX PD 08-OCT-1998.
XX XX 27-MAR-1998; 98WO-US06022.
XX XX

FT /transl_except= (pos:310..315, aa:Xaa)
 FT /note= "Xaa is unknown"
 PN CA2255109-A1.
 XX 17-JUN-1999.
 PD 17-DEC-1998: 98CA-2255109.
 XX PF 17-DEC-1997: 97CA-2255126.
 XX PR 17-DEC-1997: HSC RES & DEV LP.
 PA (HSC-) Egan SE;
 XX WPI: 2000-148082/14.
 XX P-PSDB: AAY69698.
 DR New nucleic acids encoding a murine and human Brainiac protein, useful
 PT for detecting somatic or germline DNA lesions which are responsible for
 PT developmental syndromes or diseases including cancer.
 XX
 PS Claim 5: Page 29-30; 40pp: English.
 XX
 CC This sequence represents cDNA encoding human Brainiac protein. The
 CC complete murine Brainiac cDNA sequence (AA28185) was used to screen an
 CC EST (expressed sequence tag) database to identify human Brainiac cDNA
 CC clones, which were assembled to form the human Brainiac cDNA of the
 CC present invention. Human Brainiac has significant similarity to
 CC Drosophila melanogaster Brainiac. Drosophila Brainiac and Egghead
 CC proteins regulate adhesion between epithelial cells, this activity being
 CC dependent on the presence of Notch. Drosophila Brainiac is a secreted
 CC protein, and has sequence similarities with the Drosophila Fringe
 CC proteins. Brainiac/Egghead-mediated epithelial cell adhesion is required
 CC for cell viability, cell growth regulation and cell fate specification.
 CC Wild-type or mutant forms of mammalian Brainiac proteins may therefore be
 CC used to alter epithelial cell adhesion in a mammal. Mammalian Brainiac
 CC proteins, active fragments analogues, and nucleic acids may be used to
 CC treat diseases such as cancer, psoriasis and other skin lesions, and
 CC nervous system disorders. Mammalian Brainiac nucleic acids may also be
 CC used to detect somatic or germline DNA lesions which are responsible for
 CC developmental syndromes or diseases including cancer. The mammalian
 CC Brainiac proteins and fragments or its analogues are useful as antigens
 CC in immunoassays including enzyme-linked immunosorbent assays (ELISA),
 CC radioimmunoassays (RIA) and other non-enzyme linked antibody binding
 CC assays. Non-human transgenic animals comprising nucleotide sequences
 CC encoding human Brainiac protein (AAY69698) can be used as animal models
 CC for the study of mammalian Brainiac gene function, for the screening of
 CC candidate compounds and for the evaluation of potential therapeutic
 CC interventions.
 XX
 SQ Sequence 2536 BP; 747 A; 429 C; 500 G; 803 T; 57 other;
 Query Match 72.08; Score 861.6; DB 21; Length 2536;
 Best Local Similarity 83.28; Pred. No. 5.5e-256;
 Matches 973; Conservative 0; Mismatches 194; Indels 3; Gaps 1;
 QY 30 TTGCTGGGCATCTCTGATGGCAAAATGCTTCTTATTATTGTTGGAGTCTCCAA 89
 DB 28 TTGTTGGGTATCTCTGATGGCAAAATGCTTCTTATTATTATTGTTGGAGTCTCCAA 87
 QY 90 AACAGTAGCCCAAGACAAAATGGAAGGGGAGGAGTAAATATCCGAAAGAGAGTCTCG 145
 DB 88 AGCAGTAGCCCAAGACAAAATGGAAGGGGAGGAGTAAATATATCCCAAGAGAGTCTCG 147
 QY 150 AAGCCACCCAGCACTCCCGGGGATCTGGAACAGGGAACAGGAGAGCTGAACAGGTGG 209
 DB 148 AAGATATCTACCCCTCCCGAGGCACTACTGGAACCGAGAGCAAGAGAGCTGAACCGGCAG 207
 QY 210 TACATATCCCATCTTGAACAGGTGGCCAAATACAGACAGGGGAGGTAGCCACATCTCCAAAC 269
 DB 208 TACAACCCCATCTCTGAGCATGCTATGACCAACAGAGGGGGAGGGGAGGCTCTCCAAT 267

QY 270 ACRAGTCACCTGAGCTATTGIGAACCCAGAGCTCGACGGTCTATGACAGCTGTG---ACAGAT 326
 DB 268 ATAAGTCATCTGAACCTACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTGTGGT 327
 QY 327 TTTAATAATCTGCGGACAGATTAAAGACITTTCTCTTGTATTGAGATGCGGGAATTAC 386
 DB 328 TTTAACAACCTGCGGACAGATTAAAGACTTTCTCTGTATTGAGATGCGGCAATTAT 387
 QY 387 TCGCTGCTTATAGATCAACCGAAGAAATGTGAAAGAGCCCTTCTTACTATTGGCGATA 446
 DB 388 TCACCTGCTTATAGATCAGCCGGATAGTGTGCAAGAAACCTTTCTTGTGTGGCGATT 447
 QY 447 AAGTCCCTCATTCACATTTTGCAGAGCAAGCAATTCGGGAGTCTTGGGGCCGAGAA 506
 DB 448 AAGTCCCTCACTCCACATTTTGCAGAGCAAGCAATTCGGGAATCTCTGGGGCCGAGAA 507
 QY 507 ACCAACCTAGGGACACAGACAGTAGTAGGCTCTTCTGTGGGCAAGACACCCCCAGAG 566
 DB 508 AGCAACCGAGGAACCAACGCGTGGTGGAGTCTTCTGCTGGGCCAGACACCCCCAGAG 567
 QY 567 GACAACCACTGACCTTTCGGACATGCTTAAAGTTTGAGAGTGACAGCACCAGGACATC 626
 DB 568 GACAACCACTGACCTTTCAGATATGCTGAAATTTGAGAGTCAGAAGCACCAGACATT 627
 QY 627 CTCATGTGGAACCTATAGACACATTTCTCAACCTGCTCCCTGAAGAGTGTGTTCTT 686
 DB 628 CTTATGTGGAACCTATAGACACATTTCTCAACCTGCTCTGAAGAGTGTGTTCTTCTC 687
 QY 687 AGTGGGTGAGCACTTCTCTCCAGCGCAGAGATTGCTTCTTCAAGGCGCATGATGACGTG 746
 DB 688 AGTGGGTGAGTCTCTCTCCAGCAGACATGAGTTGTTTCTCAAGGCGCATGATGAT 747
 QY 747 TTTGTGAACCCCATCATCTTAAATTAATTAATGCTTGAATAGCTTATCCAGAGCAAGCCAA 806
 DB 748 TTTGTGAACCCCATCATCTTAAATTAATTAATGCTTGAATAGCTTATCCAGAGCAAGCCAA 807
 QY 807 GACTTGTTCATAGTGTGATCCCAATGCTGGGCTTCCAGGGGATAGAAACCTGAAG 866
 DB 808 GATCTCTTCATAGTGTGATGATCCCAATGCTGGGACCTCATCGGATAGAAAGCTGAAG 867
 QY 867 TACTACATCCAGAGTCTTCTTACACCGGCTCTACCCACCTGATGCGGGGGTGTGGA 926
 DB 868 TACTACATCCAGAGTGTGTTTACTTGGCTCTTACCCACCTATGCGAGGGGAGGGGG 927
 QY 927 TTCTGTACTCCGCGCCCTTGGCTTGGAGCTGTACAGTGGGACTAGCCGGTCCATCTC 986
 DB 928 TTCTGTACTCCGCGCCCTGAGGCTGTACCATATCACTGACCAAGGTCATCTC 987
 QY 987 TACCTATATGATGTGTTATACGGGAATGTGCTTCAGAACTGGGCTTGTTCAGAG 1046
 DB 988 TACCCATGTAGAGTGTATCTGGAATGTGCTTCAGAACTCGGCTCGTTCAGAG 1047
 QY 1047 AAGCAAAAGCTTCAGACATTTGATTTGAAGAGAAAATAGAAAAATATTTGTTC 1106
 DB 1048 AAACACAAAGCTTCAGACATTTGATTTGAAGAGAAAATAGAAAAATATACATCTCTC 1107
 QY 1107 TATATAGACCTTAATGTAGTACATAGCAGAAAACCTCAAGAGATGATTGATATCTG 1166
 DB 1108 TATGTAGATCTGTAGTGTAGTACATAGTAAACCTCAAGAGATGATTGATATTTG 1167
 QY 1167 CAGTTGCAAGCTCTAAATTTAAATGCTGA 1196
 DB 1168 CAGTTGCAAGCTCTAAATTTAAATGCTGA 1197
 RESULT 15
 ID AAH14972
 XX AAH14972 standard; cDNA: 2442 BP.
 AC AAH14972;
 XX DT 26-JUN-2001 (first entry)
 XX

DE Human cDNA sequence SEQ ID NO:12897.
XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
KW Homo sapiens.
OS EPI074617-A2.
PN 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX Claim 8: SEQ ID 12897; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC of an oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
SQ Sequence 2442 BP: 717 A: 433 C: 503 G: 789 T; 0 other;
Query Match 71.0%; Score 848.6; DB 22; Length: 2442;
Best Local Similarity 84.6%; Pred. No. 5.9e-252;
Matches 953; Conservative 0; Mismatches 174; Indels 0; Gaps 0;
OY 70 TGATTGTGGAGTCTCCAAAACAGTAGCCACAGCAAAAATGAAAGGCGAGGTAATAA 129
DB 1 TTATTATGGAAGTCTCCAAAGCAGTAGCCACAGAAAATGAAAGGCGAGGTAATAA 60
OY 130 TCCCGAAGAGAGTCTTGGAGCCACCCAGCACTCCCGGGCATACTGGAGCAGGGAAC 189
DB 61 TACCCAAAGAGAGTCTTGGAGATATCTACCCCTCCCGAGGATCTACTGGAAACCGAGGC 120
OY 190 AGGAGAGAGTGAACAGGTGGTACATCCCATCTTGGACAGGTTGGCCATCAGACAGGGG 249
DB 121 AAGAGAGAGTGAACCGGAGTGAACCCCATCTCTGAGCATGCTTGACCAACCGAGCGGG 180

OY 250 AGCTAGCCACATCTCCAAACACAAAGTCACCTGAGCTATTGTGAACACAGACTCGAGCGTCA 309
DB 181 AGCGGGCAGGCTCTCCAATATAAGCCATCTGAACCTACTGCGAAGCTGACCTGAGGGTCA 240
OY 310 TGACAGCTGTGACAGATTTTAATAATCTCCGCGACAGATTTAAAGACATTTCTCTGTAT 369
DB 241 CGTCGGTGGTTACGGGTTTTAACAACTTCCCGACAGATTTAAAGACATTTCTGTGTAT 300
OY 370 TGAGATGCCGGAATTAATCTCGTCTTATAGATCAACCGAAGAAATGTGCAAGAGCCCT 429
DB 301 TGAGATGCCGCAATTTATCACTGCTTATAGATCAGCCGATTAAGTGTGCAAGAAACCTT 360
OY 430 TCTTACTATTGGCGATAAAGTCCCTCAITCCACATTTTCCGAGAGGCAAGCAATTCGG 489
DB 361 TCTGTGTGCGGATTAAGTCCCTCACTCCACATTTTCCGAGAGGCAAGCAATTCGG 420
OY 490 AGTCTTGGGCGGAGAAACCAAGTAGGGAACACAGACAGTGTAGGGTCTTCTCTGTGG 549
DB 421 AATCTTGGGCGCAAGAAAGCAACGCGAGGCAACCAACGCTGCTGCGAGTCTTCTCTGTGG 480
OY 550 GCAAGACACCCCGAGAGGACACCCCTGACCTTTCGACATGCTTAAGTTGAGAGTG 609
DB 481 GCCAGACACCCCGAGAGGACACCCCTGACCTTTCAGATATGCTGAAATTTGAGAGTG 540
OY 610 ACAAGCACAGGACATCCTCATGTGGAATATAGACACACATTTCTCAACCTGTCCCTGA 669
DB 541 AGAAGCACAGACATTTATGTGGAATATAGACACACATTTCTCAACTTGTCTCTGA 600
OY 670 AGAAGTGTGTTTCTTAGTGGGTGAGCACTTCTCTGTCAGACGACAGATTTGTCTTCA 729
DB 601 AGAAGTGTGTTTCTCAGTGGGTAACTTCTCTGCGGACACACTGAGTTGTTTCA 660
OY 730 AGGCGATGATGAGTGTGTTGTAACACCCATCCTTAATTACTTGAATAGTAT 789
DB 661 AGGCGATGAGTGTGTTGTAACACCCATCCTGAATGATGATGATGAT 720
OY 790 CCAAGCAAGCAAGCAAGTCTTTCATAGTGTGATGATGATGATGATGATGATGATGATGAT 849
DB 721 CCAAGCAAGCAAGCAAGTCTTTCATAGTGTGATGATGATGATGATGATGATGATGATGAT 780
OY 850 GGGATAAGAACTGAACTACTACATCCAGAGTCTTCTACACCGGCTCTACCCACCGT 909
DB 781 GGGATAAGAACTGAACTACTACATCCAGAGTGTGTTTACTCTGGCCTCTACCCACCGT 840
OY 910 ATGCGGGGTGTGGATTCCTTACTTCGGGCCCCCTTGCCTTGAGGCTGTACGTGGA 969
DB 841 ATGCGGGGTGTGGATTCCTTACTTCGGGCCCCCTTGCCTTGAGGCTGTACCATATCA 900
OY 970 CTAGCGGGTCCATCTCTACCCCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1029
DB 901 CTAGCGGGTCCATCTCTACCCCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 960
OY 1030 TGGGCTTGTTCAGAGAGCAAGAGGCTTCAGGACATTTGATATTTGAAGAGAAATA 1089
DB 961 TGGGCTTGTTCAGAGAGCAAGAGGCTTCAGGACATTTGATATTCAGAGAGAAATA 1020
OY 1090 AGAAATATTTGTTCTTATATAGACCTAATGTACTACATACGACAGAAACCTCAGAGA 1149
DB 1021 AAAATACATCTGCTCTCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
OY 1150 TGATTGATATCTGCTCTCAGTTGCAAAAGTCTTAATTTAAATGCTGA 1196
DB 1081 TGATTGATATTTGCTCTCAGTTCCAGAGTCTCAATTTAAATGCTAA 1127

Search completed: October 20, 2003, 14:45:55
Job time : 343.33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2003, 14:27:23 : Search time 2702.64 Seconds
(without alignments)
10755.472 Million cell updates/sec

Title: US-09-804-357b-1
Perfect score: 1196
Sequence: 1 agatgagtgtgggcgtcga.....gtcttaatttaaatgctga 1196

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match: 0%
Maximum Match: 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estro:*

7: em_estro:*

8: em_hic:*

9: gb_estl:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pin:*

20: em_gss_vit:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gssl:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	542	45.3	587	10	BE623172 up49a10.x
2	533.4	44.6	547	14	CD560301
3	529	44.2	589	9	AW495241
4	521.8	43.6	832	9	AU137528

5	521	43.6	592	10	BF467111
6	517	43.2	517	9	AA119132
7	517	43.2	743	14	CB554366
8	499.6	41.8	1201	9	AL553132
9	488	40.8	505	9	AA204363
10	483	40.4	692	10	BG106243
11	479.2	40.1	609	10	BE100018
12	474	39.6	876	13	BX370083
13	466.2	39.0	488	10	BE632125
14	456.8	38.2	476	9	AA986646
15	456.4	38.2	1053	12	BM557603
16	455.4	38.1	157	9	AA733556
17	441	36.9	501	10	BE685687
18	440.2	36.8	664	14	BY724171
19	440	36.8	489	10	BG062994
20	440	36.8	686	14	BY754595
21	433	36.2	655	14	BY753386
22	432.2	36.1	659	10	BG741357
23	417	34.9	543	9	AW530344
24	416.4	34.8	477	10	BE944633
25	412.8	34.5	558	12	BM537435
26	408.4	34.1	410	10	BF469505
27	403.6	33.7	680	14	CB457941
28	401.4	33.6	964	13	B0893660
29	400	33.4	490	9	AI408668
30	388.8	32.5	945	13	B0369620
31	379.6	31.7	610	10	BB619419
32	378.2	31.6	484	9	AI407754
33	375.4	31.4	392	9	AA929654
34	371	31.0	430	10	BE946424
35	369.8	30.9	593	10	AW958548
36	369	30.9	632	10	BG721682
37	363.8	30.4	865	13	B0427178
38	361	30.2	719	12	B1157193
39	360	30.1	372	9	AA209930
40	360	30.1	373	10	BE631971
41	359	30.0	701	9	AV724042
42	344.4	28.8	405	9	AW456055
43	337	28.2	444	9	AI070816
44	334	27.9	383	9	AI112158
45	333.8	27.9	813	12	BI664280

ALIGNMENTS

RESULT 1
BE623172
LOCUS up49a10.x1 Soares.mouse.MMGB.bcell Mus musculus cDNA clone
DEFINITION IMAGE:2749434 3' Similar to TR:054904 054904 UDP-GAL:BETAGLCNAC
BETA 1.3-GALACTOSYLTRANSFERASE-1, mRNA sequence.

ACCESSION BE623172.1 GI:9903578

VERSION BE623172.1

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 587)

AUTHORS NCI-CGAP

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

JOURNAL Tumor Gene Index

COMMENT Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgaaps@mail.nih.gov

This clone is available royalty-free through LILNL; contact the

IMAGE Consortium (info@image.lilnl.gov) for further information.

MGI:1035806

Possible reversed clone: similarity on wrong strand

High quality sequence stop: 452.

Location/Qualifiers

1..587

source

source


```
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:2749434"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares_mouse_WNGB_bcell"
/organ="germinal B-cell; Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGGCGGCGGCGCTGTTTCTTTTCTTTTCTTTT
T 3'] double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized; constructed by Bento Soares and
M.Fatima Bonaldo."
BASE COUNT 175 a 131 c 152 g 126 t 3 others
ORIGIN
Query Match 45.3%; Score 542; DB 10; Length 587;
Best Local Similarity 99.4%; Pred. No. 1.8e-130; Indels 0; Gaps 0;
Matches 542; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AGATGAGTGTGGGCGTCCGAAGAGTCAAGTTCGTGGCATCTCTGATGATGCCAAATGTCT 60
DB 43 AGATGAGTGTGGGCGTCCGAAGAGTCAAGTTCGTGGCATCTCTGATGATGCCAAATGTCT 102
QY 61 TCATTATTGATGTGGAAGTCTCCAAAACAGTACCCCAAGACACAAAATGGAAAGGAG 120
DB 103 TCATTATTGATGTGGAAGTCTCCAAAACAGTACCCCAAGACACAAAATGGAAAGGAG 162
QY 121 GAGTAATAATCCGAAGAGAGTCTTGAAGCCACCCAGCAGTCCCGGGCATCTGGA 180
DB 163 GAGTAATAATCCGAAGAGAGTCTTGAAGCCACCCAGCAGTCCCGGGCATCTGGA 222
QY 181 ACAGGAACAGGAGAGCTGAACAGTGGTACAAATCCCATCTTGAACAGGTTGCCAATC 240
DB 223 ACAGGAACAGGAGAGCTGAACAGTGGTACAAATCCCATCTTGAACAGGTTGCCAATC 282
QY 241 AGACAGGGAGCTAGCCATCTCCAAACAGTCACTGAGTATTGTGAACAGACT 300
DB 283 AGACAGGGAGCTAGCCATCTCCAAACAGTCACTGAGTATTGTGAACAGACT 342
QY 301 GCACGGTCATGACAGCTGTGACAGATTTAATAATCTGCCGACAGATTTAAGACTTC 360
DB 343 GCACGGTCATGACAGCTGTGACAGATTTAATAATCTGCCGACAGATTTAAGACTTC 402
QY 361 TCTTGATTTCGATGCGCGAATTACTCGCTGCTTATAGATCAACGCAAGAAATGTGCA 420
DB 403 TCTTGATTTCGATGCGCGAATTACTCGCTGCTTATAGATCAACGCAAGAAATGTGCA 462
QY 421 AGAAGCCCTTCTTACTATTGGCGATAAGTCCCTCAATCCACATTTTCCGAGAGGCAAG 480
DB 463 AGAAGCCCTTCTTACTATTGGCGATAAGTCCCTCAATCCACATTTTCCGAGAGGCAAG 522
QY 481 CAATTGGGAGCTTGGGCGGAGAAACCAAGTAGGGAACCCACAGACTAGTGGGCTCT 540
DB 523 CAATTGNGAGTCTTGGGCGGAGAAACCAAGTAGGGAACCCACAGACTAGTGGGCTCT 582
QY 541 TCCTG 545
DB 583 TCCTG 587
RESULT 2
CD560301 547 bp mRNA linear EST 11-JUN-2003
LOCUS B0410D02-5 NIA Mouse E6.5 Whole Embryo cDNA Library (Long) Mus
DEFINITION musculus cDNA clone NIA:B0410D02 IMAGE:30439621 5', mRNA sequence.
CD560301
ACCESSION CD560301.1 GI:31604194
VERSION EST.
KEYWORDS Mus musculus (house mouse)
```

```
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 547)
AUTHORS Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
TITLE Construction of long-transcript enriched cDNA libraries from
submicrogram amounts of total RNAs by a universal PCR amplification
method
JOURNAL Genome Res. 11 (9), 1553-1558 (2001)
MEDLINE 21429098
PUBMED 11544199
COMMENT Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: B0410 row: D column: 02
Seq primer: M13 Reverse
High quality sequence stop: 547
POLYA-No. Location/Qualifiers
1..547
/organism="Mus musculus"
/mol_type="mRNA"
/strain="G57BL/6J"
/db_xref="niaEST:B0410D02-5"
/db_xref="taxon:10090"
/clone="NIA:B0410D02 IMAGE:30439621"
/tissue_type="E6.5 whole embryo"
/dev_stage="whole embryo including extraembryonic tissues
at 6.5-days postcoitum"
/lab_host="DH10B"
/clone_lib="NIA Mouse E6.5 Whole Embryo cDNA Library (Long)
)"
/notes="Vector: pCMV-SPORT6 (Invitrogen); Site_1: SalI;
Site_2: NotI; Mouse cDNA project by the Laboratory of
Genetics, National Institute on Aging (NIA), Intramural
Research Program, NIH (http://igsun.grc.nia.nih.gov/GDNA).
This is a long-transcript enriched cDNA library (Ref.
Genome Res. 11: 1553-1558 (2001). (PMID: 11544199)). Total
RNAs were extracted from a pool of 7 embryos at 6.5-days
postcoitum. Double-stranded cDNAs were synthesized with an
oligo(dT) primer [Invitrogen:
5'-GGACTAGTCTAGTCGCGAGCGGCCCTTTTCTTTTCTTTT-3']
from 0.53 ug of total RNA, treated with T4 DNA polymerase,
and purified by ethanol-precipitation. The cDNAs were
ligated to Lope-linker L-Sal4, purified by
phenol/chloroform, and separated from free linkers by
Centricon 100. Then, the cDNAs were amplified by
long-range high fidelity PCR using Ex Taq polymerase
(Takara) with a primer Sal4-S. The products were purified
by phenol/chloroform and Centricon 100. The cDNAs were
digested with SalI and NotI enzymes and cloned into
SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E.
coli host was transformed with the ligation mixture by the
standard chemical method. The average insert size is about
2.3kb. The library was constructed by Yulan Piao."
BASE COUNT 157 a 128 c 127 g 135 t
ORIGIN
Query Match 44.6%; Score 533.4; DB 14; Length 547;
Best Local Similarity 99.6%; Pred. No. 3.2e-128;
Matches 545; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 328 TTAATAATCTCGCGACAGATTTAAAGACTTCTCTTGTATTGATGATGCGGAATFAC 387
DB 1 TTAATAATCTCGCGACAGATTTAAAGACTTCTCTTGTATTGATGATGCGGAATFAC 60
QY 388 C-GCTGCTTATAGATCAACCGAAGAAATGTCGAAGAGCCCTTCTTACTATTGGCGATA 446
DB 61 CAGTGCCTTATAGATCAACCGAAGAAATGTCGAAGAGCCCTTCTTACTATTGGCGATA 120
QY 447 AAGTCCCTCATCCACATTTTGGCCAGAGCAATTCGGGAGTCTTGGGCGCGGAG 506
```

```
Db 121 AGTCCCTCAATCCACATTTGCGAAGGCAAGCAATTCGGAGTCTTGGGGCCGAGAA 180
Qy 507 ACCAAGCTAGGAGACAGACAGTAGTGAGGGTCTTCTGCTGGCAAGACACCCGCCAGAG 566
Db 181 ACCAAGCTAGGAGACAGACAGTAGTGAGGGTCTTCTGCTGGCAAGACACCCGCCAGAG 240
Qy 567 GACAACACCCCTGACCTTTCCGACATGCTTAAGTTTGAGAGTGACAAGCACCAGACATC 626
Db 241 GACAACACCCCTGACCTTTCCGACATGCTTAAGTTTGAGAGTGACAAGCACCAGACATC 300
Qy 627 CTCATGTGGAACTATAGACACATCTTCAACCTGTCCCTCAGGAAGTGTCTTCTT 686
Db 301 CTCATGTGGAACTATAGACACATCTTCAACCTGTCCCTCAGGAAGTGTCTTCTT 360
Qy 687 AGTGGGTGAGCACTTCTCTGTCAGACGAGAGTTTGTCTTCAAGGGCGATGACGCTG 746
Db 361 AGTGGGTGAGCACTTCTCTGTCAGACGAGAGTTTGTCTTCAAGGGCGATGACGCTG 420
Qy 747 TTGTGACACCATCACATCTTACTTACTTGAATAGCTTATCCAGAGCAAGGCCAAA 806
Db 421 TTGTGACACCATCACATCTTACTTACTTGAATAGCTTATCCAGAGCAAGGCCAAA 480
Qy 807 GACTTGTTCATAGTGCAGTGATCCACATGCTGGCCCTCACCGGGGATAGAACTGAAG 866
Db 481 GACTTGTTCATAGTGCAGTGATCCACATGCTGGCCCTCACCGGGGATAGAACTGAAG 540
Qy 867 TACTACA 873
Db 541 TACTACA 547
```

```
RESULT 3
AW495241 589 bp mRNA linear EST 24-FFB-2000
LOCUS UT-M-BH3-aus-b-04-0-UI.s1 N:R:BMAP_M_S4 Mus musculus cDNA clone
DEFINITION UT-M-BH3-aus-b-04-0-UI 3', mRNA sequence.
ACCESSION AW495241
VERSION AW495241.1 GI:7065522
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (Bases 1 to 589)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996);
97044477
8889548
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
Oligo-dT track not found, Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA library Preparation: M.B.
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENETICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA-No. Location/Qualifiers
1. 589
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
```

FEATURES
source

```
/clone="UI-M-BH3-aus-b-04-0-UI"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NIH_BMAP_M_S4"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not 1; Site_2: Eco RI; The
NIH_BMAP_M_S4 library is a subtracted library of a series,
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M_S4,
NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S1,
NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
(NIH_BMAP_M_S4) was constructed as follows: PCR amplified
cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
libraries in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library)
was purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
NIH_BMAP_M_S4 library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LIB=NIH_BMAP_M_S4
TAG_TISSUE=hippocampus
TAG_SEQ=TTCGA*
BASE COUNT 174 a 136 c 154 g 124 t 1 others
ORIGIN
Query Match 44.2%; Score 529; DB 9; Length 589;
Best Local Similarity 99.8%; Pred. No. 4.6e-127;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 AGATGAGTGTGGGGCTCGAAGAGTCAAGTTGCTGGGCATCCTCGATGATGGCAAAATGTCT 60
Db 60 AGATGAGTGTGGGGCTCGAAGAGTCAAGTTGCTGGGCATCCTCGATGATGGCAAAATGTCT 119
Qy 61 TCATTTATTGATGTGAGAGTCTCCAAAACAGTAGCCCAAGACAAAATGGAAGGGAG 120
Db 120 TCATTTATTGATGTGAGAGTCTCCAAAACAGTAGCCCAAGACAAAATGGAAGGGAG 179
Qy 121 GAGTAATAATCCGAAAGAGAAGTTCTGGAAGCCACCCAGCACTCCCGGGGCATCTGGA 180
Db 180 GAGTAATAATCCGAAAGAGAAGTTCTGGAAGCCACCCAGCACTCCCGGGGCATCTGGA 239
Qy 191 ACAGGGAACAGGAGAGCTGAACAGGTGGTACAAATCCCATCTTGAACAGGGTGGCCATC 240
Db 240 ACAGGGAACAGGAGAGCTGAACAGGTGGTACAAATCCCATCTTGAACAGGGTGGCCATC 299
Qy 241 AGACAGGGGAGCTAGCCACATCTCCAAACACAACTCACTGAGCTATGTGAACCCAGACT 300
Db 300 AGACAGGGGAGCTAGCCACATCTCCAAACACAACTCACTGAGCTATGTGAACCCAGACT 359
Qy 301 CGAGGGTCATGACAGCTGTGACAGATTTTAAATCTGCCGGACAGATTTAAAGACTTTC 360
Db 360 CGAGGGTCATGACAGCTGTGACAGATTTTAAATCTGCCGGACAGATTTAAAGACTTTC 419
Qy 361 TCTTGTATTGAGATGGCGGAATTACTCGCTGTATTAGATCAACCGAAGAAATGTGCAA 420
Db 420 TCTTGTATTGAGATGGCGGAATTACTCGCTGTATTAGATCAACCGAAGAAATGTGCAA 479
Qy 421 AGAAGCCCTTCTTACTATTGGCGGATAAAGTCCCTCATTTCCACATTTTGGCAGAGGCAAG 480
Db 480 AGAAGCCCTTCTTACTATTGGCGGATAAAGTCCCTCATTTCCACATTTTGGCAGAGGCAAG 539
Qy 481 CAATTGGGGAGTCTTGGGGCCGAGAAACCAACAGTAGGGAAACACAGACAGTA 530
```

```

|||||
540 CAATTCGGAGTCTTGGGCGGAGAACCAAGCTANGGAACACAGACAGTA 589

RESULT 4
AUI37528 832 bp mRNA linear EST 02-AUG-2002
LOCUS AUI37528 PLACE1 Homo sapiens cDNA clone PLACE1006678 5', mRNA
DEFINITION
ACCESSION AUI37528
VERSION AUI37528
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 832)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
Location/Qualifiers
1..832
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1006678"
/tissue_type="placenta"
/clone_lib="PLACE1"
/notes="Vector: pME18SFL3"
BASE COUNT 248 a 193 c 194 g 194 t 3 others
ORIGIN

Query Match 43.6%; Score 521.8; DB 9; Length 832;
Best Local Similarity 80.9%; Pred. No. 3.9e-125;
Matches 656; Conservative 0; Mismatches 149; Indels 6; Gaps 4;

QY 77 GGAAGTCTCCAAAACAGTAGCCCAAGACAAAATGGAAGGGAGGAGTAAATATCCCGNA 136
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 9 GAAAGTCTCCAAAAGCAGTAGCCCAAGAAAATGGAAGGGAGGAGTAAATATACCCAA 68
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 137 AGAGAAGTCTCGAAGCCACCCAGCAGTCTCCCGGGCATCTGGAACAGGAGGAGGAGAA 196
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 69 AGAGAAGTCTCGAAGATCTACCCCTCCGAGGCGATCTGGAACGAGAGCAGAGNA 128
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 197 GCTGAACAGGTGTACAAATCCCAATCTTGAAACAGGGTGGCCAAATCAGACAGGGAGGTAGC 256
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 129 GCTGAACCCGCGAGTACAAACCCCATCTGTAGCATGCTGACCAACAGCGGGGGGCGGG 188
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 257 CACATCTCCAAACAGCAGTACCTGAGCTATTGTGAACGAGATCGACGGTCTATGACACC 316
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 189 CAGGCTCTCAATATAAGCCATCTGAACTACTCGGAACCTGACCTGAGGGTCACGTCGT 249
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 317 TGTGCAGAGATTTAAATATCTGGCGGACAGATTAAAGACTTTCTCTTATTTAGATG 376
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 249 GGTACGGGTTTAAACAACTTGGCGGACAGATTAAAGACTTTCTCTGTATTGAGATG 308
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 377 CCGGAATTAATCGCTGCTTATAGATCAACGAGAAATGCAAGAACCCCTTCTTACT 436
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 309 CCGCAATTATTAATCGCTGCTTATAGATCAACGAGAAATGCTGCAAGAACCCCTTCTTGT 368
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 437 ATTGGCGATAAAGTCCCTCATTTCCACATTTTCCAGAGGCAAGCAATTCGGAGTCTTG 496
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 369 GCTGGCGATTAAAGTCCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTG 428
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 497 GGGCCGAGAAACCAAGTAGGGAACGACAGACAGTACAGGGTCTCTCTCTTGGCGAAGC 556
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 429 GGGCCGAGAAACCAAGTAGGGAACGAGGGAACCAACGGTGGTGGAGTCTCTCTCTG 488
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 557 ACCCCAGAGGAGCAACACCCCTGACCTTTCCGACATGCTTAAAGTTTGAGAGTGAACA 616
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 489 ACCCCAGAGGAGCAACACCCCTGACCTTTCCGACATGCTTAAAGTTTGAGAGTGAACA 548
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 517 CCAGGACATCTCATGTGGAACATATAGACACATCTCTCAACCTGCTCCGCAAGAAAGT 676
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 549 CCAAGACATCTTATGTGGAACATATAGACACATCTCTCAACCTGCTCCGCAAGAAAGT 608
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 677 GCTGTTTCTTAGTGGGTGAGCACATCTCTGTCACAGCAGAGTGTGTCTTCAAGGGGA 736
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 609 GCTGTTTCTCAAGTGGTAACTTCTCTGCGCAGACACTGAGTTGTTTCAAGGGGA 668
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 737 TGATGAGGTGTTGTGAACACCATCATCATCTCTTAATTAATTT--GAATAGCTTATCC 794
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 569 TGACGATGTTTTGTGAACACCATCATCATCTCTGGAATTTACTTTGAATAGTTTATCC 728
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 795 AGCAAGCCCAAGACTT--GTTTATAGTGTGATGATCCCAATGCTGGG--CCTCACCGG 851
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 729 ACCAAGCCCAAGACTTCTTCAAAAGTTTGTATGATGATCCCAATGCTGGGACCTCATCG 788
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 852 GATAAGNAAGTGAAG-TACTACATCCAGAA 881
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 789 GATAAGNAAGTGAAGTTACTACTATCCANAA 819
Db | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
BF467111 592 bp mRNA linear EST 04-DEC-2000
LOCUS UI-M-CGOp-bre-g-04-0-UI-s1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone
DEFINITION UI-M-CGOp-bre-g-04-0-UI 3', mRNA sequence.
ACCESSION BF467111
VERSION BF467111.1 GI:11536294
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 592)
AUTHORS Bonaudo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
Oligo-dT track not found, Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab clone distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENETICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA-No. Location/Qualifiers
1..592
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"

```



```

|||||
157 GACAGCAGGACATCTTCATGTGGAACATATAGAGACACATTTCAACCTGTCCCTG 98
|||||
669 AAGGAAGTGTCTTCTTAGTGGGTGAGCACTTCCTGTCAGAGCAGAGTTTGCTTC 728
|||||
97 AAGGAAGTGTCTTCTTAGTGGGTGAGCACTTCCTGTCAGAGCAGAGTTTGCTTC 38
|||||
729 AAGGGCGATGATGACGTGTTTGTGAACACCCATCACA 765
|||||
37 AAGGGCGATGATGACGTGTTTGTGAACACCCATCACA 1
|||||

RESULT 7
CB554366
LOCUS CB554366 743 bp mRNA linear EST 01-JUN-2003
DEFINITION MMSP0063_H02 MMSP Macaca mulatta cDNA, mRNA sequence.
ACCESSION CB554366
VERSION CB554366.1 GI:31303561
KEYWORDS EST.
SOURCE Macaca mulatta (rhesus monkey)
ORGANISM Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
1 (bases 1 to 743)
Katze,M.G., Bungarner,R., Korth,M., Feldman,R., Amjadi,M. and
Holzman,T.
Expressed sequence tags from Rhesus macaque spleen
Contact: Holzman T
Katze Lab
University of Washington
Box 358070, Seattle, WA 98195-8070, USA
Tel: 206 732 6156
Fax: 206 732 6055
Email: ted@locke.hs.washington.edu
Similar to GenBank entry BC030579 BC030579 Homo sapiens,
UDP-GlucNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 1,
clone MGC:26071 IMAGE:4828158, mRNA, complete cds. 5/2002
Plate: MMSP0063 row: H column: 02.
FEATURES
Location/Qualifiers
1..743
/organism="Macaca mulatta"
/mol_type="mRNA"
/db_xref="taxon:9544"
/sex="male"
/cell_type="monocytes"
/dev stage="adult"
/clone.lib="MMSP"
/note="Organ: spleen"
BASE COUNT 200 a 182 c 175 g 182 t
ORIGIN
Query Match 43.2% Score 517; DB 14; Length 743;
Best Local Similarity 82.9% Pred. No. 6,6e-124;
Matches 614; Conservative 0; Mismatches 125; Indels 2; Gaps 2;
QY 185 GGACAGAGAGAGTGTGAACAGGTGTGTACATCCATCTTGACAGAGGTGGCCATCAGAC 244
|||||
DB 1 GGCACAGATGATGAACAGCGGAGTACACACCCCATCTGAGCATGTAGCCCAACCCAGAC 60
|||||
QY 245 AGGGGAGTACCCACATCTCCAAACACAGTCACTGAGCTATTGTGAACAGACCTGCAC 304
|||||
DB 61 GGGGAGGCGGGGAGGCTCTCCAAATATAGCCATCTGAACATTTGTGAGCCTGACCTGAG 120
|||||
QY 305 GGTGATGACAGTGTGACAGATTTTAATAATCTGCCGACAGATTTAAAGACATTTCTT 364
|||||
DB 121 GGTGATGACAGTGTGACAGATTTTAATAATCTGCCGACAGATTTAAAGACATTTCTGCT 180
|||||
QY 365 GTATTGAGATGCCGGGAATTAATCTGCTGCTTATAGATCAACCCAGAAATATGCAAGAA 424
|||||
DB 181 GTATCTGAGATGCCGGGAATTAATCTGCTGCTTATAGATCAACCCAGGGAATGTCAGAA 240
|||||

```

```

QY 425 GCCCTTCTTACTATGGCGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAGCAAT 484
|||||
DB 241 ACCCTTCTTGTGCTGGCGATTAAAGTCCCTCACTCCACATTTTCCAGAGGCAAGCAAT 300
|||||
QY 485 TCGGGAGTCTTGGGCGGAGAAACCAACGTTAGGGAACCCAGACAGTAGTGGGGTCTTCT 544
|||||
DB 301 TCGGGAAATCTTGGGCGGAGAAACCAACGTTAGGGAACCCAGACAGTAGTGGGGTCTTCT 360
|||||
QY 545 GTTGGGCAAGACACACCCCGAGAGGAAACCAACCCCTTTCGGACATGCTTAAAGTTGA 604
|||||
DB 361 GCTGGCGAGACACCCCGAGAGGAAACCAACCCCTTTCAGATATGCTGAAATTTGA 420
|||||
QY 605 GAGTGACAGCACCAGGACATCTCATGTGGAACTATAGAGACACATTCCTCAACCTGTC 664
|||||
DB 421 GAGTGACAGCACCAGGACATCTCATGTGGAACTATAGAGACACATTCCTCAACCTGTC 480
|||||
QY 665 CTTGAAGGAAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 724
|||||
DB 481 TCTGAAGGAAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 540
|||||
QY 725 CTTCAAGGGCGATGATGACGTGTTTGTGAACACCCCATCACATCTTAATTCATGATAG 784
|||||
DB 541 TTTCAAGGGCGATGATGATGTTTGTGAACACCCCATCACATCTTGAATTTACTTGAATAG 600
|||||
QY 785 CTTATCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 844
|||||
DB 601 TTTATCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 660
|||||
QY 845 TCACCGGGGATAAGAACTGAAGTACTACATCCAGAAAGTCTTCTACACCGGGGTACCC 904
|||||
DB 661 TCATCGGGGATAAGAACTGAAGTACTACATCCAGAAAGTCTTCTACACCGGGGTACCC 718
|||||
QY 905 ACCGTATGTCGGGGGTGGTGG 925
|||||
DB 719 ACCGTATGTCGGGGGTGGGGG 739
|||||

AL553132 1201 bp mRNA linear EST 31-MAY-2003
AL553132 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
Clone CS001074YK14 5-PRIME, mRNA sequence.
ACCESSION AL553132
VERSION AL553132.2 GI:31274946
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
L.I.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 15, 2001 this sequence version replaced gi:12892683.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 1713.f For
more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS001074BF07QPLcluster=1713.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS001074BF07QPL.
FEATURES
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS001074YK14"
/tissue_type="PLACENTA COT 25-NORMALIZED"

```



```

Db      265  GACATCCTCATGTGGAACTATAGACAGACATCTTCAACCTGCTCCTGAAGAAAGTGCTG 206
Qy      681  TTCTTTAGTGGGTGAGCATTCTCTGTCAGACGACGAGTTTGTCTTCAAGGGCGATGAT 740
      |||||
Db      205  TTCTTTAGTGGGTGAGCATTCTCTGTCAGACGACGAGTTTGTCTTCAAGGGCGATGAT 146
Qy      741  GAGGTGTTGTGAACACCCATCATCTCTTAATTACTTGAATAGCTTATCCAGAGCAAA 800
      |||||
Db      145  GAGGTGTTGTGAACACCCATCATCTCTTAATTACTTGAATAGCTTATCCAGAGCAAA 86
Qy      801  GCCAAGACATTTGTATAGTGCAGTGCATCCACAATGCTGGCCCTCAC -CGGGA-AAAGAA 859
      |||||
Db      85  GCCAAGACATTTGTATAGTGCAGTGCATCCACAATGCTGGCCCTCACCTCGGGAAGAA 26
Qy      860  ACTGAAGTACTACATCCAG 879
      |||||
Db      25  ACTGAAGTACTACATCCAG 6

RESULT 10
LOCUS   BG106243
DEFINITION 602289589f1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4384414 5',
      mRNA sequence.
ACCESSION BG106243
VERSION   BG106243.1 GI:12600089
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
      Email: cgabbs-remail.nir.gov
      Tissue Procurement: Louis Staudt, M.D., Ph.D.
      cDNA Library Preparation: Life Technologies, Inc.
      cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
      DNA Sequencing by: Incyte Genomics, Inc.
      Clone distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LLNL at:
      http://image.llnl.gov
      Plate: LLAM10064 row: a column: 23
      High quality sequence stop: 687.
      Location/Qualifiers
          1..692
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:4384414"
              /tissue_type="lymphoma, cell line"
              /lab_host="DH10B (phage-resistant)"
              /clone_lib="NIH_MGC_85"
              /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
              Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
              Average insert size 1.867 kb. Library enriched for
              full-length clones and constructed by Life Technologies.
              Note: this is a NIH_MGC Library."
BASE COUNT 186 a 166 c 167 g 172 t 1 others
ORIGIN
Query Match 40.4%; Score 483; DB 10; Length 692;
Best Local Similarity 84.6%; Pred. No. 4.9e-115;
Matches 566; Conservative 0; Mismatches 100; Indels 3; Caps 2;
Qy      265  CAACACAGTCACCTCAGCTATTGTCAACACGACTCGACGGTCATGACAGCTGTCAGAC 324
      |||||
Db      23  CCATATAGGCACTCAAGTACTCGCAACCTGACCTGAGGGTCAGCTCGGTGGTTACGG 82
Qy      325  ATTTTAATAATCTGCCGGACAGATTTAAAGACTTTCTTGTATTGAGATGCCGAAT 384
      |||||

```

```

Db      83  GTTTTAACAACTGCCGGACAGATTTAAAGACTTTTCTGCTGTATTGAGATGCCCAAT 142
Qy      385  ACTGCTGCTTATAGATCAACCGAAGAAATGTGCAAGAAAGCCCTTCTTACTTATGGGCA 444
      |||||
Db      143  ATTCACCTGCTTATAGATCAGCGGATAAGTGTGCAAGAAACCTTCTTGTGCTGGCGCA 202
Qy      445  TAAAGTCCCTCATTCACCATTTTGCAGAGGCAAGCAATTCGGGAGTCTTGGGGCGCGAG 504
      |||||
Db      203  TTAAGTCCCTCACTCCACAT -TTTGCAGAGGCAAGCAATTCGGGAATCCTGGGCGCAAG 261
Qy      505  AAACCAACGCTAGGGAACCAACAGACAGTASTGAGGAGTCTTCTGTTGGSCAGACACCCCGAG 564
      |||||
Db      262  AAAGCAAGCGAGGAGCAACCAACCGTGTGCGAGTCTTCTGCTGGGCCAGACACCCCGAG 321
Qy      565  AGGACAACCAACCTGACCTTTGGGACATGCTTAAGTTTGGAGAGTGACAAGACACCAAGACA 624
      |||||
Db      322  AGGACAACCAACCCCGACCTTTTCAGATATGCTGAATTTTGAGAGTGAGAAAGACCAAGACA 381
Qy      625  TCCTCATGTGGAACATATAGAGACACATCTTCAACCTGTCCCTGAGGAAGTGCCTGTTTC 684
      |||||
Db      382  TTCTTATGTGGAACATACAGACACATTTCTCAACTTGTCTGGAAGAGTGTCTGTTTC 441
Qy      685  ITAGGTGGGTGAGCAGCTTCTGTCAGAGCAGAGTTTGTCTTCAAGGGCGAGTATGACGG 744
      |||||
Db      442  TCAGGTGGGTAACTACTTCTGTCAGACACTGAGTTTCTTTCAGAGGCGAGTACGAGT 501
Qy      745  TGTGTTGGAACACCCATCACATCTTAAATTACTTTGTAATAGCTTATCCAGAGCAAGACA 804
      |||||
Db      502  TTTTGTGGAACACCCATCACATCTGTAATTTGTAATAGTTTATCCAGAGCAAGACA 561
Qy      805  AAGACTTTGTCATAGTGTGAGTGATGATCCACAAATGCTGGGCTCACCGGATAAGAACTGA 864
      |||||
Db      562  AAGATCTTTCATAGTGTGATGTGATCCACAAATGCTGGACCTCATCGGGATAAGAACTGA 621
Qy      865  AGTACTACATCCCA -GAAGTCTTCTACACCGCGCTTACCCACCGTATGCCGGGGGTGG 922
      |||||
Db      622  AGTACTACATCCCAAGGAAGTTGTTTACTCTGGGCTCAACCCACTATGCGAGGGGAGG 661
Qy      923  TGGATTCTCT 931
      |||||
Db      682  GGGGGTCT 690

RESULT 11
LOCUS   BE100018
DEFINITION UI-R-BJ1-atr-g-05-0-UI.s1 UI-R-BJ1 Rattus norvegicus cDNA clone
      UI-R-BJ1-atr-g-05-0-UI 3', mRNA sequence.
ACCESSION BE100018
VERSION   BE100018.1 GI:8491903
KEYWORDS EST.
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
      Rattus.
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
      discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
      Coordinated Laboratory for Computational Genomics
      University of Iowa
      375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
      Tel: 319 335 8250
      Fax: 319 335 9565
      Email: bento-soares@uiowa.edu
      Oligo-dT track not found, Not 1 site shown in beginning of sequence
      is likely internal to the message. cDNA Library Preparation: M.B.
      Soares Lab Clone distribution: clones will be available through

```

```

Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA-No.      Location/Qualifiers
1. .609
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/lab_host="PH10B (Life Technologies)"
/clone_lib="UI-R-BJ1"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker: Site_1: Not I; Site_2: Eco RI; The UI-R-BJ1
library is a subtracted library derived from the following
tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV
canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc,
AV canal at 15 dpc, ventricle at 13 dpc, and adult heart.
For a detailed description of the library from which this
clone was derived, please visit our web site at
ratest.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_SEQ=None found"
BASE COUNT      169 a 139 c 165 g 134 t 2 others
ORIGIN
Query Match      40.1%; Score 479.2; DB 10; Length 609;
Best Local Similarity 91.8%; Pred. No. 4.7e-114;
Matches 505; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 1 AGATGAGTGTGGGCGTCGAAGAGTCAAGTTCGTGGGCATCTCGATGATGGCAATGTCT 60
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 60 AATGAGTGTGGGCGTCGAAGAGTAAAGTTCGTGGGCATCTCGATGATGGCAATGTCT 119
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 TCATTATTGATGTGGAGTCTCCAAACAGTAGCCCAAGACAAACAAATGGAAGGGAG 120
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 120 TCATCTATTGATGTGGAGTCTCCAAACAGTAGCCCAAGACAAAGATGCAAGGGGG 179
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 121 GAGTATAATACCGAAGGAGAGTCTGGAAGCCACCCAGCACTCCCGGGCATACTGA 180
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 180 GAGTATAATACCGAAGGAGAGTCTGGAAGCCACCCAGCACTCCCGGGCATACTGA 239
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 181 ACAGGGAACAGGAGAGTGTGAACAGTGTGTACAAATCCCATCTTGAACAGGTGGCCAAATC 240
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 240 ATAGGGAACAGGAGAGTGTGAACAGTGTGTACAAATCCCATCTTGAACAGGTGGCCAAATC 299
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 241 AGACAGGGAGCTAGCCACATCTCCAAACAGTCAACCTGAGCTATTGTGAACAGACT 300
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 300 AGACGGGGATCTGTTCACATCTCCAAACAGTCAACCTGAGCTATTGTGAACAGACT 359
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 301 CGACGCTCATGACAGCTGTGACAGATTTTAATTAATCTGCGGACAGATTTAAGACTTC 360
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 360 CGACAGTCAATGACAGCTGTGACAGATTTTAATTAATCTGCGGACAGATTTAAGACTTC 419
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 361 TCTGTATTGATGATCGCGAATTACTTCGCTGCTTATAGATCAACCGAAGAAATGTGCA 420
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 420 TCTGTATTGATGATCGCGAATTACTTCGCTGCTTATAGATCAACCGAAGAAATGTGCA 479
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 421 AGAAGCCCTTCTTACATTATGGGATTAAGTCCCTCATTCACATTTTGGCAGAGGCAAG 480
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 480 AGAAGCCCTTCTTACATTATGGGATTAAGTCCCTCATTCACATTTTGGCAGAGGCAAG 539
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 481 CAATTCTGGGAGCTCTGGGGCCGAGAACCAACCTAGGGAACACAGACTAGTAGGGGTCT 540
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 540 CAATTCTGGGAGCTCTGGGGCCGAGAACCAACCTAGGGAACACAGACTAGTAGGGGTCT 599
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 541 TCCTGTGGG 550
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 600 TCCTTTTGGG 609
DB 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

BX370083 876 bp mRNA linear EST 08-MAY-2003
 BX370083 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
 clone CS0D1074YK14 5-PRIME, mRNA sequence.
 BX370083
 EST. BX370083.1 GI:30449879
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 876)
 Li.W.B., Gruber C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 1713.f For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0BAG036ZC11_CS03443_1&cluster=1713.f.
 Contact : Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0BAG036ZC11_CS03443_1.
 Location/Qualifiers
 1. 876
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0D1074YK14"
 /tissue_type="PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo (3P)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."
 BASE COUNT 248 a 172 c 196 g 258 t 2 others
 ORIGIN

Query Match 39.6%; Score 474; DB 13; Length 876;
 Best Local Similarity 86.7%; Pred. No. 1.2e-112;
 Matches 534; Conservative 0; Mismatches 80; Indels 2; Gaps 1;

QY 583 TTTCGGACATGCTTAAGTTTTCAGAGTGACAGACACAGG--ACATCTCATGTGGAAC 640
 DB 3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 3 TTGAGGATGCTGCTGAATTTTGAGAGTGAGAGACCAAGNNACATTTCTTATGTGGAAC 62
 QY 641 TAGAGACACATCTTCAACCTGTCCTCAAGGAAGTGTCTTCTAGTGGGTGAGCAC 700
 DB 63 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 63 CAGAGACACATTTCTTCAACCTGTCCTCAAGGAAGTGTCTTCTAGTGGGTGAGTAC 122
 QY 701 TTCTGTCCAGACGACAGAGTTTGTCTTCAAGGGGATGATGACCTGTTTGTGAACACCA 760
 DB 123 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 123 TTCTGTCCAGACACTGAGTTTGTTCAGGGGATGACCATGTTTGTGGAACACCA 182
 QY 761 TCACATCTTAAATTAATTAAGTCTTATCCAGAGCAAGCCAAAGACATTTGTTTCATAGG 820
 DB 183 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 183 TCACATCTGAAATTAATTAAGTCTTATCCAGACCAAGCCAAAGATCTCTTCATAGG 242
 QY 821 TCACGTGATCCACAATGCTGGCCCTCACCGGATGAAGAACTGAAGTACTACATCCACGA 880
 DB 243 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 243 TGATGTGATCCACAATGCTGGCCCTCACCGGATGAAGAACTGAAGTACTACATCCACGA 302
 QY 881 AGTCTTCTACACCGCGCTTACCCACCGTATGCGGGGGTGTGGATTCTCTGTTACTCGG 940
 DB 303 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 303 AGTTGTTTACTCTGGCCCTTACCCACCGTATGCGGGGGTGTGGATTCTCTGTTACTCGG 362
 QY 941 CCCCCCTTGGAGCTGTACAGTCGCACTAGCCGGTCCCATCTCTACCTCTATTGATGA 1000
 DB 363 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 363 CCACCTGGCCCTTGAGGCTGTACCATATCACTACACAGGTCCCATCTCTACCCCATGTGA 422


```

QY 1001 TGTATTACGGGAATGTCCTTCAGAAACTGGGCTTGTCTCCAGAGACGACAAAGGCTT 1060
Db 423 CTTTATCTAGTAATGTCCTTCAGAAACTGGGCTTGTCTCCAGAGACGACAAAGGCTT 482
QY 1061 CAGGACATTTGATATTGAAGACAAAATAAGAAAAATATTGTTCCCTATATAGACCTAAT 1120
Db 483 CAGGACATTTGATATTGAAGACAAAATAAGAAAAATATTGTTCCCTATATAGATCTGAT 542
QY 1121 GTTAGTACATACGACAAACCTCAAGAGATGATTGATATCTGGTCTCAGTTGCAAAAGTCC 1180
Db 543 GTTAGTACATACGACAAACCTCAAGAGATGATTGATATTTGGTCTCAGTTGCAAGATCC 602
QY 1181 TAAATTAAATGCTGA 1196
Db 603 TCAATTAAATGCTAA 618

RESULT 13
BE632125/c
LOCUS
DEFINITION
us57d09.y1 Soares_thymus_2NDMT Mus musculus cDNA clone
IMAGE:3376049 5' similar to TR:Q54904 Q54904 UDP-GAL:BETAGLCNAC
BETA 1.3-GALACTOSYLTRANSFERASE-1. ;, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 488)
REFERENCE
AUTHORS
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished
CONTACT: Robert Strausberg, Ph.D.
Email: cgaaps-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1083653
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 455.
FEATURES
Location/Qualifiers
1..488
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3376049"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/clone_lib="Soares_thymus_2NDMT"
/polylinker="Vector: pT73D-Pac (Pharmacia) with a modified
Not I; Site 1; Site 2; Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAGTGGAGCGCGGCTTTTATTTTATTTTATTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT 136 a 112 c 108 g 132 t
ORIGIN
Query Match 39.0%; Score 466.2; DB 10; Length 488;
Best Local Similarity 99.4%; Pred. No. 1.1e-110;
Matches 468; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 669 AAGGAAGTCTGTTCTTAGTGGGTGAGCAGCTTCTCTGCCAGACGACAGATTGTTCTTC 728
Db 488 AAGGAAGTCTGTTCTTAGTGGGTGAGCAGCTTCTCTGCCAGACGACAGATTGTTCTTC 429
QY 729 AAGGGCGATGATCAGCTGTTTGTGAACACCCATCATCTTAATTACTTGAATAGCTTA 788
Db 428 AAGGGCGATGATGACGTGTTTGTGAACACCCATCATCTTAATTACTTGAATAGCTTA 369
QY 789 TCCAAGAGCAAGCCAAAGACTTGTTCATAGTGCATGATCCCAATGCTGGGCTCAC 848
Db 368 TCCAAGAGCAAGCCAAAGACTTGTTCATAGTGCATGATCCCAATGCTGGGCTCAC 309
QY 849 CGGGATAAGAACTGAAGTACTACATCCAGAGAGTCTTCTACACGGCGTCTACCCACGG 908
Db 308 CGGGATAAGAACTGAAGTACTACATCCAGAGAGTCTTCTACACGGCGTCTACCCACGG 249
QY 909 TATGCCGGGGGTGGTGCATTCTCTACTCCGGCCCCCTTGCCTTGAGGCTGTACAGTGG 968
Db 248 TATGCCGGGGGTGGTGCATTCTCTACTCCGGCCCCCTTGCCTTGAGGCTGTACAGTGG 189
QY 969 ACTAGCCGGGTCCATCTCTACCTTATGATGATGTTTATACGGGATGTCCTTCAGAAA 1028
Db 188 ACTAGCCGGGTCCATCTCTACCTTATGATGATGTTTATACGGGATGTCCTTCAGAAA 129
QY 1029 CTGGGCTTGTCCAGAGAGACAAAGGCTTCAGGACATTTGATATTGAAGAGAAAAT 1088
Db 128 CTGGGCTTGTCCAGAGAGACAAAGGCTTCAGGACATTTGATATTGAAGAGAAAAT 69
QY 1089 AAGAAAAATATTGTTCCCTATATAGACCTTAATGTTAGTACATAGCAGAAAA 1139
Db 68 AAGAAAAATATTGTTCCCTATATAGACCTTAATGTTAGTACATAGCAGAAAA 18

RESULT 14
AA986646 476 bp mRNA linear EST 28-MAY-1998
LOCUS
uc82d09.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
IMAGE:1432145 5' similar to TR:Q24157 Q24157 NEUROGENIC SECRETED
SIGNALLING PROTEIN. ;, mRNA sequence.
AA986646
VERSION
KEYWORDS
SOURCE
AA986646.1 GI:3168300
Mus musculus (house mouse)
ORGANISM
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 476)
REFERENCE
AUTHORS
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE
The WashU-HMI Mouse EST Project
JOURNAL
Unpublished
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:916213
Seq primer: primer name ambiguous
High quality sequence stop: 474.
FEATURES
Location/Qualifiers
1..476
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1432145"

```

```
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Sugano mouse kidney mKia"
/notes="Organ: Kidney; Vector: pME18S-FL3; Site_1: DraIII
(CACTGTGTC); Site_2: DraIII (CAGCATGTC); 1st strand cDNA
was primed with an oligo(dT) primer
(ATGTGGCTTTTITTTTITTTT); double-stranded cDNA was
ligated to a DraIII adaptor (TGTGGCTTACGG), digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGTC, 3' site CAGCATGTC). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTCTGCTCTATAAAGCTGG and 3' end
primer CGACCTGAGCTCGAGACA."
BASE COUNT 122 a 125 c 121 g 108 t
ORIGIN
Query Match 38.2%; Score 456.8; DB 9; Length 476;
Best Local Similarity 99.4%; Pred. No. 3.1e-108;
Matches 469; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 477 CAAGCAATTCGGGAGTCTTGGGGCCGAGAAACCAACGTTAGGAAACAGACAGTAGTGAGG 536
Db 1 CAAGCAATTCGGGAGTCTTGGGGCCGAGAAACCAACGTTAGGAAACAGACAGTAGTGAGG 60
QY 537 GTCTTCCTGTGGGCAAGACACCCCGAGAGGACACCCCTGACCTTTCCGACATGCTT 596
Db 61 GTCTTCCTGTGGGCAAGACACCCCGAGAGGACACCCCTGACCTTTCCGACATGCTT 120
QY 597 AAGTTTTCAGTTCACAGCAGGACATCCTCATGTGGAACATATAGACACACATTCCTC 656
Db 121 AAGTTTTCAGTTCACAGCAGGACATCCTCATGTGGAACATATAGACACACATTCCTC 180
QY 657 AACCTGTCCTGAAGGAAGTGTGTTCTTAGTGGGTGAGCACTTCCTGTCAGACGCA 716
Db 181 AACCTGTCCTGAAGGAAGTGTGTTCTTAGTGGGTGAGCACTTCCTGTCAGACGCA 240
QY 717 GAGTTTGTCTTCAGGGCGATGATGACGTGTTTGGAACACCCATCAGATCCTTAATAC 776
Db 241 GAGTTTGTCTTCAGGGCGATGATGACGTGTTTGGAACACCCATCAGATCCTTAATAC 300
QY 777 TTGAATAGCTTATCCAAAGAGCAAGCCAAAGACTTGTTCATAGGTGACGTGATCCACAT 836
Db 301 TTGAATAGCTTATCCAAAGAGCAAGCCAAAGACTTGTTCATAGGTGACGTGATCCACAT 360
QY 837 GCTGGSCCTACCGGGATAGAAACTGAAGTACTATACATCCAGAACTCTTCTACACCGGC 896
Db 361 GCTGGSCCTACCGGGATAGAAACTGAAGTACTATACATCCAGAACTCTTCTACACCGGC 420
QY 897 GTCTACCCACCGTATCCCGGGGT-GGTGGATTCTGTATCICCGGCCCTT 947
Db 421 GTCTACCCACCGTATCCCGGGGTGGGGGATTCTGTATCICCGGCCCTT 472
RESULT 15
BM557603
LOCUS
DEFINITION
5', mRNA sequence.
ACCESSION
BM557603
VERSION
BM557603.1 GI:18799721
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1053)
NIH-MGC http://mgc.nci.nih.gov/
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
```

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12754 row: c column: 21
High quality sequence stop: 547.

FEATURES

Location/Qualifiers

1..1053

source

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5739980"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH-MGC_88"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 Kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH-MGC Library."

BASE COUNT 288 a 240 c 219 g 306 t

ORIGIN

Query Match 38.2%; Score 456.4; DB 12; Length 1053;
Best Local Similarity 87.5%; Pred. No. 4.9e-108;
Matches 499; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 627 CTCATGTGGAACCTATAGAGACACATCTTCAACCTGTCCTCGAAGAAAGTGCIGTTTCT 686
Db 1 CTTATGTGGAACCTATAGAGACACATCTTCAACTTGTCTCGAAGAAAGTGCIGTTTCT 60
QY 687 AGTGGGTGAGCAGCTTCCTGTCAGACCGGCTTGTCTTCAAGGGGATGATCAGCTG 746
Db 61 AGTGGGTGAGTACTTCTGTCAGACACACTGAGTTGTTTCAAGGGGATGAGCATGTT 120
QY 747 TTTGTGAACACCCATCAGATCCTTAAATTACTTGAATAGCTTATCAAGAGCAAGCCAAA 806
Db 121 TTTGTGAACACCCATCAGATCCTTGAATTACTTGAATAGTTTATCAAGACCAAGCCAAA 180
QY 807 GACTTGTTCATAGGTGAGCTGATCCACAATGCTGGGCTCACCAGGATAGAAAGTGAAG 866
Db 181 GATCTCTTCATAGGTGATGTGATCCACAATGCTGGACCTCATCGGGATAGAAAGCTGAAG 240
QY 867 TACTACATCCAGAAAGTCTTACACCGGCTCTACCCACCGTATGCCGGGGTGGTGA 926
Db 241 TACTACATCCAGAAAGTGTGTTTACTCTGGCCTCTACCCACCGTATGCCAGGGGGGGGG 300
QY 927 TTCCTGTACTCGGGCCCTTTCCTTTGAGGCTGTACAGTGGGAGTACGCCGGGTGCATCTC 986
Db 301 TTCCTGTACTCGGGCCACCTGSCCTGAGGCTGTACCATATCATCTGACCAAGTCCATCTC 360
QY 987 TACCCATTGATGATGTTTATACGGGAATGTCCTTACAGAACTGGGCTTGTTCACAG 1046
Db 361 TACCCATTGATGAGCTTTATATCTGGAATGTGCTTTCAGAACTCGGCTCGTTCACAG 420
QY 1047 AAGCAAAAGCTTCAGGACATTTATATTCAGAGAAATTAAGAAATATTTGTTC 1106
Db 421 AAACAAAGGCTTCAGGACATTTGATATCAGGAGAAACAAAATAAGATATGCTGCC 480
QY 1107 TATATAGACCTTAATGTTAGTACATAGCAGAAACCTCAAGAGATGATGATATCTGGTCT 1166
Db 481 TATGTAGATCTGATGTTAGTACATAGTAAACCTCAAGAGATGATGATATTTGGTCT 540
QY 1167 CAGTTGCAAGTCTTAATTTAAATGCTGA 1196
Db 541 CAGTTGCAAGTCTTAATTTAAATGCTGA 570

Search completed: October 20, 2003, 19:37:12
Job time : 2719.64 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw mode:

Run on: October 20, 2003, 14:46:23 : Search time 76.2177 Seconds
(without alignments)
6926.136 Million cell updates/sec

Title: US-09-804-357b-1
Perfect score: 1196
Sequence: 1 agatgagtggggcgctga.....gtcctaatttaaaagctga 1196

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq:.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	159.4	14.2	1134	4	US-09-482-180A-3
2	163	13.6	1191	4	US-09-459-133-3
3	156.8	13.1	1167	4	US-09-459-133-14
4	136.6	11.4	1420	4	US-09-482-180A-1
5	130.2	10.9	1532	4	US-09-459-133-1
6	123	10.3	1446	4	US-09-459-133-12
7	44.6	3.7	2095	4	US-09-996-243-208
8	43.2	3.6	7218	1	US-08-232-463-14
9	41.8	3.5	1434	2	US-09-055-097-2
10	35.2	2.9	1680	5	PCT-US93-06748-1
11	35.2	2.9	2601	3	US-09-117-217-9
12	35.2	2.9	2601	3	US-09-117-217-11
13	35.2	2.9	2601	3	US-09-117-217-13
14	35.2	2.9	2601	4	US-09-735-487-7
15	35.2	2.9	2601	4	US-09-735-487-9
16	35.2	2.9	2601	4	US-09-735-487-11
17	35.2	2.9	2601	4	US-09-735-487-13
18	35.2	2.9	2601	4	US-09-743-357-16
19	35.2	2.9	3033	2	US-07-743-357-19
20	35.2	2.9	3856	2	US-07-743-357-20
21	35.2	2.9	4113	2	US-07-743-357-21
22	35.2	2.9	4307	4	US-09-552-950-1
23	35.2	2.9	5362	3	US-08-463-210-5
24	35.2	2.9	5610	4	US-09-262-537-57
25	35.2	2.9	8932	3	US-09-124-900-1
26	35.2	2.9	8933	3	US-08-463-210-4
27	35.2	2.9	8933	3	US-08-463-210-4

ALIGNMENTS

RESULT 1

US-09-482-180A-3
Sequence 3, Application US/09482180A

Patent No. 6361985

GENERAL INFORMATION:

APPLICANT: Conklin, Darrell C.

APPLICANT: Yamamoto, Gayle

APPLICANT: Gao, Zeren

APPLICANT: Whitmore, Theodore E.

APPLICANT: Jaspers, Stephen

TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING

FILE REFERENCE: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOG, ZN5SP6

CURRENT APPLICATION NUMBER: US/09/482,180A

CURRENT FILING DATE: 2000-01-12

PRIOR APPLICATION NUMBER: 60/115,721

PRIOR FILING DATE: 1999-01-12

NUMBER OF SEQ ID NOS: 20

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 3

LENGTH: 1134

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

NAME/KEY: degenerate sequence

LOCATION: (1)...(1134)

OTHER INFORMATION: n is any nucleotide

NAME/KEY: misc_feature

LOCATION: (1)...(1134)

OTHER INFORMATION: n = A,T,C or G

US-09-482-180A-3

Query Match 14.2%; Score 169.4; DB 4; Length 1134;

Best local Similarity 34.7%; Pred. No. 6.9e-45;

Matches 308; Conservative 140; Mismatches 416; Indels 24; Gaps 5;

Qy	312	ACAGCTGTGCACAGATTTTAATATCGCCGACAGATTTAAAGACTTCTCTTGTTATTG	371
	11	11	11
Db	241	ACNGTWNWSNGCNWSNVTNNSNTNCCNWSNCAVMGNHTYTTNACNTAYMGN	300
	11	11	11
Qy	372	AGATCGCGGAATCTACTCGCTTATAGATCAACCGAAGAAATGTCCAAAGAGCCCTTC	431
	11	11	11
Db	301	CAYTGYMNAATVTVWSNATHVTNTNG---ARCCNWSNGGNTGYWSNAARGAYACNTTY	357
	11	11	11
Qy	432	TTACTATTGGCATAAAGTCCCTCATTCACATTTTCCAGAGGCAACCAATTCGGGAG	491
	11	11	11
Db	358	YTNHTYNTGNCNATHAARWNSNCARCCNGCNGCATGNTNGARMGNMNGCNGCNATMGNWSN	417
	11	11	11
Qy	492	TCTTGGGCGCGAGAAACCAACGTAGGGAACACAGTAGTCAGGCTCTTCTCTGTTGGGC	551

Db 418 ACNTGGGNGM-----NGTNGNGNGTGGGCGNMGNGNMGNCARYTNAARYNTNGN 468
QY 552 AACACACCCCCAGAGACACACCCTGACCTTTCGGACATGCTTAAGTTGAGATGAC 611
Db 469 TTYTYNTYNGNGTNGNGNWSNGNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 528
QY 612 AACACACAGGACATCTCATGTGGAACTATAGACACACATCTTCAACCTGCTCCCTGAAG 671
Db 529 GARTTYGAYATHVNTCAATGGGAYTYACNGARGAYTYTYAAYTYNACNYTNAAR 588
QY 672 GAAGTCTGTTCTTAGTGGGTGAGCACTTCTGTCCAGACCCAGAGTTTGTCTCAAG 731
Db 589 GARYTHCAYYTNGARNGTGGGTNGTNGCNGCTGYCCNCCNCCNCCNCCNCCNCCN 648
QY 732 GCGGATCATGACGTGTGTGGAAACCCATCATCTTAAITACTTGAATACCTTAACC 791
Db 649 GNGAYGAYGAYGNTTYTNCAYGTNCCNCCNCCNCCNCCNCCNCCNCCNCCN 703
QY 792 AAGAGCAAGCCAAAGACATGTTTCATAGGTGAGCTGATCCACAAATGCTGGGCTACCCG 851
Db 704 -GGAYCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 762
QY 852 GATAAGAACTGAAGTACTACATCCAGAGATCTTCTACA---CCGGCGTCTACCCACCG 908
Db 763 AATACNARGTNAARTATTTATTCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 822
QY 909 TATCGCGGGGTGGTGGATTCCTGACTCCGCCCCCTTGCCTTGAGGTGTACAGTGG 968
Db 823 TAYCGNGGNGGNGGNTAYGTNATGWSNMGNCNCCNCCNCCNCCNCCNCCNCCN 882
QY 969 ACTAGCGGGTCCATCTACCCATGATGATGATGATGATGATGATGATGATGATG 1028
Db 883 ATGARGAYGNGARNTYNTNWSNATHTGAYGAYGAYGAYGAYGAYGAYGAYGAY 942
QY 1029 CTGGGCTGTTCAGAGAGACAAAGGCTTCAGGACATTTGATATT--GAAGAGAA 1085
Db 943 YTNNGNYTNSNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 1002
QY 1086 AATAAGAAAAATTTGTTCTATATAGACCTTAATCTTAGTACATAGCAGAAAACTCAA 1145
Db 1003 GAYCCNYTNGAYCCNTGYTNTAYTMGNGGNYTNTYNTCTNCCNCCNCCNCCNCCN 1062
QY 1146 GAGATGATGATATCTGCTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 1193
Db 1063 GARATGGACNATGTGGCGNYTNGTNGACNGAYGARGGNYTNAARTGY 1110

RESULT 2

US-09-459-133-3
: Sequence 3, Application US/09459133
: Patent No. 6416988
: GENERAL INFORMATION:
: APPLICANT: Conklin, Darrell C.
: APPLICANT: Yamamoto, Gayle
: APPLICANT: Jaspers, Stephen R.
: APPLICANT: Gao, Zeren
: TITLE OF INVENTION: BETA-1.3-GALACTOSYLTRANSFERASE HOMOLOGS
: FILE REFERENCE: 98-77
: CURRENT APPLICATION NUMBER: US/09/459,133
: CURRENT FILING DATE: 1999-12-10
: PRIOR FILING DATE: 1999-12-10
: NUMBER OF SEQ ID NOS: 20
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 3
: LENGTH: 1191
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Degenerate sequence
: FEATURE:
: NAME/KEY: misc_feature

: LOCATION: (1)...(1191)
: OTHER INFORMATION: n - A,T,C or G
US-09-459-133-3

Query Match 13.6%; Score 163; DB 4; Length 1191;
Best Local Similarity 34.0%; Pred. No. 8.7e-43;
Matches 261; Conservative 110; Mismatches 379; Indels 18; Gaps 2;

QY 426 CCCTTCTTACTATTGGCGATAAAGTCCCTCATCTTCACATTTTGGCAGAGCAAGCAAT 485
Db 442 CCAATAYTYNTYNTGCGTNAARWSNGNCCNCCNCCNCCNCCNCCNCCNCCNCCN 501
QY 486 CCGGAGTCTTGGGCGGAGAAACCAACCTAGGAAACAGACAGTAGTGGAGGTCTCTCG 545
Db 502 MNGARACNTGGGWSNCCNCC-----NCCNGNATHMGNYTNTYNTY 546
QY 546 TTGGCAAGACACCCCGAGAGGACAAACCCCTGACCTTTCGGACATCTTCAACCTGCC 605
Db 547 YTYNTYNGNWSNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 606
QY 506 AGTGAAGACAGCAGCAGACATCTTCATGTGGAACTATAGACACATCTTCAACCTGCC 665
Db 607 WSNMNGNNTAYWSNGAYTYNTYNTGCGAYTYTYTNGAYGTNCCNCCNCCNCCNCCN 666
QY 666 CTGAAGAGAGTCTGTTCTTAGGTGGTGAGCACTTCTCTCCAGAGCAGAGTTGTC 725
Db 667 YTNAAARGAYTYNTYNTYNTGCGAYTYTYTNGCGAYTYTYTNGCGAYTYTYTNGCG 726
QY 726 TTCAAGGGGATGATGAGCTGTTCTGAACCCCATCATCTTAAITACTTGAATGAC 785
Db 727 YTNMNGNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 786
QY 786 TTATCAAGAGCAAAAGCAAGACTTGTTCATAGTACGTCAGTCCACAAATGCTGGGCT 845
Db 787 YTNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 846
QY 846 CACCGGGATGAAGAACTGAAGTACTACATCCAGAGTCTTCTACACCGGCTTACCCA 905
Db 847 YTNMNAARCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 906
QY 906 CCGTATCGCGGGGTGGTGGATCTCTGCTACCTCCGCCCCCTGCTTGGGCTGAGT 965
Db 907 GNTAYGNCNWSNGGNGGNTAYGTNATHGCGNMGNYTNGCNCNCCNCCNCCNCCNCCN 966
QY 966 GCGATAGCAGGCTCATCTCTACCTATGATGATGATGATGATGATGATGATGATG 1025
Db 967 GCGNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 1026
QY 1026 AAAGTGGGCTTGTTCAGAGAGCAAAAGGCTTCAGGACATTTGATTAAGAGAGAA 1085
Db 1027 GCTYNTGNGNYTNGTNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 1086
QY 1086 AATAAGAAAAATTTGTTCTCATATAGACCTTAAGTGTAGTACATAGCAGAAAACTCAA 1145
Db 1087 ACNCGNGAYCA---YTGCGNTTYMGNAAYTYNTYNTGNGMNGNCCNCCNCCNCCNCCN 1143
QY 1146 GAGATGATGATATCTGCTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 1193
Db 1144 GGNWSNATHMGNYTNTGGAARCAATYTCNARGAYCCNMGNYTNCARTGY 1191

RESULT 3

US-09-459-133-14.
: Sequence 14, Application US/09459133
: Patent No. 6416988
: GENERAL INFORMATION:
: APPLICANT: Conklin, Darrell C.
: APPLICANT: Yamamoto, Gayle
: APPLICANT: Jaspers, Stephen R.
: APPLICANT: Gao, Zeren
: TITLE OF INVENTION: BETA-1.3-GALACTOSYLTRANSFERASE HOMOLOGS
: FILE REFERENCE: 98-77
: CURRENT APPLICATION NUMBER: US/09/459,133


```
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07

Query Match      3.7%:  Score 44.6;  DB 4;  Length 2095;
Best Local Similarity 46.0%;  Pred. No. 0.00054;
Matches 189;  Conservative 0;  Mismatches 219;  Indels 3;  Gaps 1;

QY 383 TTACTCGCTGCTTATAGATCAACCGAAGAAATGTGCAAGAGCAATTCGGAGTCTTTGGGCGG 442
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 510 TCACCTTCACACTTCGAGAGCATTCAAACCTGCTCTCATCAAAATCCATTTCTGGTCACT 569
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 443 GATAAAGTCCCTCATTCACATTTTGCAGAGGCAAGCAATTCGGAGTCTTTGGGCGG 502
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 570 GGTGACCTCCACCCCTTCAGATGTGAAAGCGCAGCGCCATTAGAGTTACTTTGGGGTGA 529
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 503 AGAAACCAACGTAGGGAACCAAGACAGTAGTGAGGGTCTTCTCTGTGGGCAAGACACCC 562
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 630 AAAAAAGTCTTGGTGGGATATGAGTTCTTACATTTTCTTATTAGGCCAAGAGCTGA 689
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 563 AGAGGACAACCCCTGACCTTCGAGACATGCTTAAGTTTGAGAGTGACAAAGCACCAGGA 622
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 690 AAGGAAGACAAATGTTGGCATTGTCC---TTAGAGGATGAACACCTTCTTTATGGTGA 746
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 623 CATCTCATGTGGAACTATAGAGACACATCTTCAACCTGTCCCTGAAGGAAGTGTGT 682
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 747 CATATCGACAAGATTTTITAGACACATATATAACCTGACCTTGAACACCATTATGGC 806
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 683 TCTTAGTGGGTGAGCACCTCTCTCCAGACGACAGTGTGTTCTTCAAGGGCGAIGATGA 742
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 807 ATTCAGGTGGTAACTAGTGTTCGCCCAATGCCAAGTAGCTAATGAAGACAGACACTGA 866
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 743 CGTGTCTTGAACACCCATCACATCCTTAATTACTTGAATAGCTTATCCAA 793
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 867 TGTTTTCATCAATACTAGCAATTTAGTGAAGTATCTTTTAACCTAAACCA 917
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 8
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOXLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMNU
```

```
/
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (703)836-9300
/ TELEFAX: (703)683-4109
/ TELEX: 899149
/ INFORMATION FOR SEQ ID NO: 14:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 7218 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ CLONE: ptzpt-Fls
/ US-08-232-463-14

Query Match          3.6%; Score 43.2; DB 1; Length 7218;
Best Local Similarity 7.1%; Pred.No. 0.0033;
Matches 24; Conservative 172; Mismatches 140; Indels 0; Gaps 0;

QY 1 AGATGAGTGTGGGCGTCGAAGAGTCAAGTTGCTGGGCATCTCTGATGATGGCAATGTCT 60
Db 1352 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1293
QY 61 TCATTATTGTTGTTGAGTCTCAAAAACAGTAGCCCAAGACAAAATGAAAGGGAG 120
Db 1292 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1233
QY 121 GAGTAATAATCCGGAAGAGAACTTCTGGAAGCCACCCAGCACTCCCGGGGATCTGCA 180
Db 1232 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1173
QY 181 ACAGGGAACAGGAGAGTGAACAGTGGTGAACATCCCATTTGAACAGGGTGGCAATC 240
Db 1172 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1113
QY 241 AGACAGGGAGTAGCCACATCTCCAAACACAGTCACTGAGCTATTGTGAACAGACT 300
Db 1112 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1053
QY 301 CGACGGTCAATGACAGCTGTGACAGATTTTAATAATC 336
Db 1052 CGACCTGCACCAAGTCGGAATTAATCTGTGAGC 1017

RESULT 9
US-09-055-097-2
/ Sequence 2, Application US/09055097
/ Patent No. 5955282
/ GENERAL INFORMATION:
/ APPLICANT: Hillman, Jennifer L.
/ APPLICANT: Guegler, Karl J.
/ APPLICANT: Corley, Neil C.
/ APPLICANT: Shah, Purvi
/ APPLICANT: Patterson, Chandra
/ TITLE OF INVENTION: HUMAN OXIDIZED LDL RECEPTOR
/ NUMBER OF SEQUENCES: 7
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Incyte Pharmaceuticals, Inc.
/ STREET: 3174 Porter Drive
/ CITY: Palo Alto
/ STATE: California
/ COUNTRY: USA
/ ZIP: 94304
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/055,097
/ FILING DATE: Filed Herewith
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
```

```
/
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Cerrone, Michael C.
/ REGISTRATION NUMBER: 39,132
/ REFERENCE/DOCKET NUMBER: PF-0490 US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (650) 855-0555
/ TELEFAX: (650) 845-4166
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1434 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ LIBRARY: DUODNOT02
/ CLONE: 1705085
/ US-09-055-097-2

Query Match          3.5%; Score 41.8; DB 2; Length 1434;
Best Local Similarity 48.8%; Pred.No. 0.0035;
Matches 143; Conservative 0; Mismatches 147; Indels 3; Gaps 1;

QY 474 AGGCAAGCAATTCGGGAGTCTTGGGCGGAGAAACAGTAGGAAACACAGACACTAGTG 533
Db 298 AGAAACGCCATTTCGGGCTTCGTGGGGGGGCTGCGGAGGCCCGGGGCTCAGGGTACAG 357
QY 534 AGGGTCTTCCTGTTGGGCAAGACACCCCAAGAGGACAAC---CACCCGACCTTTCCGAC 590
Db 358 AGCGTATTCTTGGTGAGAGCCGAAACGACAGCACCCCGTGTGGGGTTCCCGAGGGAGT 417
QY 591 ATGCTTAAGTTTGAGAGTGACAAGACACAGGACATCCTCATGTGGAAGTATAGACACA 650
Db 418 GACCTGGGCTCGGATGACAGCCCGAGGGGGATATCTTGCAGGGCGGCTTCCAGGACTCC 477
QY 651 TTCTTCAACCTGTCCTCTGAAGGAAGTGTCTTTCTTAGTGTGGGTGAGCAGCTTCTGTGCA 710
Db 478 TACCGCAACCTCACCTTAAGACCTCAGCGGGCTGAACCTGGCTGAGAAACACTGCCGCC 537
QY 711 GACGACAGATTGTCTTCAAGGGCGCATGATGACGTGTTTGTGAACACCCATCA 763
Db 538 ATGGCCCGATAGTCTCTCAAGACGACGATGATGTATGTATGATCAAGGTCCTGA 590

RESULT 10
PCT-US93-06748-1
/ Sequence 1, Application PC/TUS9306748
/ GENERAL INFORMATION:
/ APPLICANT: Yung-Kang Chow
/ APPLICANT: Martin S. Hirsch
/ APPLICANT: Debra P. Meiril
/ APPLICANT: Joan C. Kaplan
/ APPLICANT: Richard T. D'Aquila
/ TITLE OF INVENTION: CONVERGENT COMBINATION ANTI-VIRAL
/ TITLE OF INVENTION: THERAPY
/ NUMBER OF SEQUENCES: 1
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Fish & Richardson
/ STREET: 225 Franklin Street
/ CITY: Boston
/ STATE: Massachusetts
/ COUNTRY: U.S.A.
/ ZIP: 02110-2804
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
/ COMPUTER: IBM PS/2 Model 502 or 55SX
/ OPERATING SYSTEM: MS-DOS (Version 5.0)
/ SOFTWARE: WordPerfect (Version 5.1)
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US93/06748
/ FILING DATE: 19930719
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
```

```

; APPLICATION NUMBER: 07/916,212
; FILING DATE: 17 July 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/145001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1680
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
PCT-US93-06748-1

Query Match 2.9%; Score 35.2; DB 5; Length 1680;
Best Local Similarity 50.0%; Pred. No. 0.55; Mismatches 0; Indels 0; Gaps 0;
Matches 88; Conservative 0;

QY 657 AACCTGTCCCTGAAGGAAGTCGTCTTTCTTAGTGGGTGAGCACTTCCTGTCCAGACGCA 716
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 860 AAGCACTAACAGAAGTAATACCCTAACAGAGAAGCAGAGCTAGAAGCTGGCAGAAAACA 919
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 717 GAGTTTGCTTCAAGGGCGATGATGAGCTGTTTGTGAACACCCATCACATCCTTAATTAC 775
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 920 GAGAGATTCTAAAGAACCAAGTACATGAGTGTATTATGACCCATCAAAGACTTAATAG 979
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 777 TTGAATAGCTTATCCAGAGCAAGAGCAAGAGCTTTTCATAGTGAGCGTATCCA 832
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 980 CAGAAATACAGAAGCAGGGGCAAGGCCAATGGACATATCAAAATTTATCAAGAGCCA 1035
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
US-09-117-217-7
; Sequence 7, Application US/09117217
; Patent No. 6221578
; GENERAL INFORMATION:
; APPLICANT: de BETHUNE, Marie-Pierre
; APPLICANT: HERTOGS, Kurt
; APPLICANT: PAUWELS, Rudi
; TITLE OF INVENTION: METHOD OF MANAGING THE CHEMOTHERAPY OF PATIENTS WHO ARE
; TITLE OF INVENTION: HIV POSITIVE BASED ON THE PHENOTYPIC DRUG SENSITIVITY
; FILE REFERENCE: 1377-125P
; CURRENT APPLICATION NUMBER: US/09/117.217
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2601
; TYPE: DNA
; ORGANISM: HIV-HXB2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (J)..(492)
; OTHER INFORMATION: gag Polyprotein
;
US-09-117-217-7

Query Match 2.9%; Score 35.2; DB 3; Length 2601;
Best Local Similarity 50.0%; Pred. No. 0.73; Mismatches 0; Indels 0; Gaps 0;
Matches 88; Conservative 0;

QY 657 AACCTGTCCCTGAAGGAAGTCTGCTTTCTTAGTGGGTGAGCACTTCCTGTCCAGACGCA 716
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1609 AAGCACTAACAGAAGTAATACCCTAACAGAGAAGCAGAGCTAGAAGCTGGCAGAAAACA 1668
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 717 GAGTTTGCTTCAAGGGCGATGATGAGCTGTTTGTGAACACCCATCACATCCTTAATTAC 776
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1669 GAGAGATTCTAAAGAACCAAGTACATGAGTGTATTATGACCCATCAAAGACTTAATAG 1728
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Query Match 2.9%; Score 35.2; DB 3; Length 2601;
Best Local Similarity 50.0%; Pred. No. 0.73; Mismatches 88; Indels 0; Gaps 0;
Matches 88; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 657 AACCTGTCCTGAAGCAAGTGTCTTTCTTAGTGGTGAGCACTTCTCTGTCAGAGCGCA 716
DB 1609 AAGCACTAACAGAGTAATACCACTAACAGAGAGCAGCTAGAACTGGCAGAAACA 1668
QY 717 GAGTTTGTCTTCAAGCGCGATGATGAGTGTGTTGTAACACCCATCACATCCTTAATAC 776
DB 1669 GAGAGATTCTAAAGAAGCAAGTACATGAGTGTATATGACCCATCAAAAGACTTAATAG 1728
QY 777 TTGAATAGCTTATCCAGAGCAAGCAAGCAAGTGTCTTATAGGTGACGTGATCCA 832
DB 1729 CAGAAATACAGAGCAGGGGCAAGGCCAATGGACATATCAAAATTTATCAAGAGCCA 1784

RESULT 14
US-09-117-217-13
; Sequence 13, Application US/09117217
; Patent No. 6221578
; GENERAL INFORMATION:
; APPLICANT: de BETHUNE, Marie-Pierre
; APPLICANT: HERTOGS, Kurt
; APPLICANT: PAUWELS, Rudi
; TITLE OF INVENTION: METHOD OF MANAGING THE CHEMOTHERAPY OF PATIENTS WHO ARE
; TITLE OF INVENTION: HIV POSITIVE BASED ON THE PHENOTYPIC DRUG SENSITIVITY
; FILE REFERENCE: 1377-125P
; CURRENT APPLICATION NUMBER: US/09/117,217
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 2601
; TYPE: DNA
; ORGANISM: HIV-HXB2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (750)..(2435)
; OTHER INFORMATION: Reverse Transcriptase
US-09-117-217-13

Query Match 2.9%; Score 35.2; DB 3; Length 2601;
Best Local Similarity 50.0%; Pred. No. 0.73; Mismatches 88; Indels 0; Gaps 0;
Matches 88; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 657 AACCTGTCCTGAAGCAAGTGTCTTTCTTAGTGGTGAGCACTTCTCTGTCAGAGCGCA 716
DB 1609 AAGCACTAACAGAGTAATACCACTAACAGAGAGCAGCTAGAACTGGCAGAAACA 1668
QY 717 GAGTTTGTCTTCAAGCGCGATGATGAGTGTGTTGTAACACCCATCACATCCTTAATAC 776
DB 1669 GAGAGATTCTAAAGAAGCAAGTACATGAGTGTATATGACCCATCAAAAGACTTAATAG 1728
QY 777 TTGAATAGCTTATCCAGAGCAAGCAAGCAAGTGTCTTATAGGTGACGTGATCCA 832
DB 1729 CAGAAATACAGAGCAGGGGCAAGGCCAATGGACATATCAAAATTTATCAAGAGCCA 1784

RESULT 15
US-09-735-487-7
; Sequence 7, Application US/09735487
; Patent No. 6528251
; GENERAL INFORMATION:
; APPLICANT: de BETHUNE, Marie-Pierre
; APPLICANT: HERTOGS, Kurt
; APPLICANT: PAUWELS, Rudi
; TITLE OF INVENTION: METHOD OF MANAGING THE CHEMOTHERAPY OF PATIENTS WHO ARE
; TITLE OF INVENTION: HIV POSITIVE BASED ON THE PHENOTYPIC DRUG SENSITIVITY
; TITLE OF INVENTION: OF HUMAN HIV STRAINS
; FILE REFERENCE: 1377-125P

; CURRENT APPLICATION NUMBER: US/09/735,487
; CURRENT FILING DATE: 2000-12-14
; PRIOR APPLICATION NUMBER: 09/117,217
; PRIOR FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2601
; TYPE: DNA
; ORGANISM: HIV-HXB2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(492)
; OTHER INFORMATION: gag Polyprotein
US-09-735-487-7

Query Match 2.9%; Score 35.2; DB 4; Length 2601;
Best Local Similarity 50.0%; Pred. No. 0.73; Mismatches 88; Indels 0; Gaps 0;
Matches 88; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 657 AACCTGTCCTGAAGCAAGTGTCTTTCTTAGTGGTGAGCACTTCTCTGTCAGAGCGCA 716
DB 1609 AAGCACTAACAGAGTAATACCACTAACAGAGAGCAGCTAGAACTGGCAGAAACA 1668
QY 717 GAGTTTGTCTTCAAGCGCGATGATGAGTGTGTTGTAACACCCATCACATCCTTAATAC 776
DB 1669 GAGAGATTCTTAAAGAGCAGCTAGCTATGAGTGTATATGACCCATCAAAAGACTTAATAG 1728
QY 777 TTGAATAGCTTATCCAGAGCAAGCAAGCAAGTGTCTTATAGGTGACGTGATCCA 832
DB 1729 CAGAAATACAGAGCAGGGGCAAGGCCAATGGACATATCAAAATTTATCAAGAGCCA 1784

Search completed: October 20, 2003, 19:40:24
Job time : 83.2177 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2003, 19:41:00 ; Search time 330.002 Seconds
(without alignments)
9517.665 Million cell updates/sec

Title: US-09-804-357B-1
Perfect score: 1196
Sequence: 1 agatgagtggtggcgctga.....gtcctaatttaaatgtga 1196

Scoring table: IDENTITY_NUC
Gapop 10.0 , Capext 1.0

Searched: 1750203 seqs, 1313063994 residues

Total number of hits satisfying chosen parameters: 3500406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*

1: /cgn2.6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2.6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2.6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2.6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2.6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2.6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2.6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2.6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2.6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
10: /cgn2.6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2.6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2.6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2.6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
14: /cgn2.6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2.6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2.6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2.6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1196	100.0	1196	9	US-09-804-357-1		Sequence 1, Appli
2	1196	100.0	1196	10	US-09-804-006-1		Sequence 1, Appli
3	1192.8	99.7	1260	13	US-10-109-563-1		Sequence 1, Appli
4	1191	99.6	1191	9	US-09-804-357-3		Sequence 3, Appli
5	1191	99.6	1191	10	US-09-804-006-3		Sequence 3, Appli
6	906.4	75.8	1707	9	US-09-804-357-13		Sequence 13, Appl
7	906.4	75.8	1707	10	US-09-804-006-13		Sequence 13, Appl
8	906.4	75.8	1831	13	US-10-109-563-3		Sequence 3, Appli
9	906.4	75.8	2745	10	US-09-972-912-1		Sequence 1, Appli
10	288.8	24.1	857	10	US-09-972-912-20		Sequence 20, Appl
11	236	19.7	480	10	US-09-972-912-26		Sequence 26, Appl
12	224.6	18.8	282	10	US-09-972-912-11		Sequence 11, Appl
13	224.6	18.8	282	10	US-09-972-912-34		Sequence 34, Appl
14	223.6	18.7	303	10	US-09-972-912-31		Sequence 31, Appl
15	211.2	17.7	266	10	US-09-972-912-12		Sequence 12, Appl
16	211.2	17.7	266	10	US-09-972-912-37		Sequence 37, Appl

ALIGNMENTS

RESULT 1

US-09-804-357-1

: Sequence 1, Application US/09804357

: Patent No. US20010024808A1

: GENERAL INFORMATION:

: APPLICANT: White, David

: APPLICANT: Zhou, Jianghong

: APPLICANT: Tartaglia, Louis A.

: TITLE OF INVENTION: LEPTIN INDUCED GENES

: FILE REFERENCE: 07334/109001

: CURRENT APPLICATION NUMBER: US/09/804,357

: PRIOR FILING DATE: 2001-03-12

: PRIOR FILING DATE: 1998-11-19

: PRIOR APPLICATION NUMBER: US 60/108,379

: PRIOR FILING DATE: 1998-10-29

: PRIOR APPLICATION NUMBER: US 09/150,857

: PRIOR FILING DATE: 1998-09-10

: NUMBER OF SEQ ID NOS: 17

: SOFTWARE: FastSeq for Windows Version 3.0

: SEQ ID NO 1

: LENGTH: 1196

: TYPE: DNA

: ORGANISM: Mus musculus

US-09-804-357-1

Query Match 100.0%: Score 1196; DB 9; Length 1196;

Best Local Similarity 100.0%: Pred. No. 0;

Matches 1196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGATGAGTGTGGCGCTGGAAGAGTCAAGTTGCTGGGCATCTCTGATGCGCAAAATGCT 60

Db 1 AGATGAGTGTGGCGCTGGAAGAGTCAAGTTGCTGGGCATCTCTGATGCGCAAAATGCT 60

Qy 61 TCATTTATTTGATTTGTTGGAAGTCTCCAAAGACAGTAGCCAAAGAAAATGGAAGGAG 120

Db 61 TCATTTATTTGATTTGTTGGAAGTCTCCAAAGACAGTAGCCAAAGAAAATGGAAGGAG 120

Qy 121 GAGTAATAATCCGAAAGAGAAGTTCTGGAAGCCACCAGCACTCCCGGGGCATACTTGA 180

```
121 GAGTAATAATCCGGAAGAGAGTTCTGGAAGCCACCAGCACATCCCGGGGCATACTGGA 180
181 ACAGGGAACAGAGAGAGCTGAACAGGTGGTGAACATCCCATCTTTGAACAGGTGGGCAATC 240
181 ACAGGGAACAGAGAGAGCTGAACAGGTGGTGAACATCCCATCTTTGAACAGGTGGGCAATC 240
241 ACACAGGGGAGTACGCCACATCTCCAAACACAAAGTCACTGAGCTATTGTGAACAGACT 300
241 ACACAGGGGAGTACGCCACATCTCCAAACACAAAGTCACTGAGCTATTGTGAACAGACT 300
301 CCACGGTCATGACAGCTGTGACAGATTTTAATAATCTGCCGACACAGATTTAAAGACTTTC 360
301 CCACGGTCATGACAGCTGTGACAGATTTTAATAATCTGCCGACACAGATTTAAAGACTTTC 360
361 TCTTGTATTGAGATGCCGGAATTAATCTCGTCTTATAGATCAACCGAAGAAATGTGCA 420
361 TCTTGTATTGAGATGCCGGAATTAATCTCGTCTTATAGATCAACCGAAGAAATGTGCA 420
421 AGAAGCCCTTCTTACTATTGGGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAG 480
421 AGAAGCCCTTCTTACTATTGGGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAG 480
481 CAATTCCGGAGTCTTGGGCGGAGAAACCAAGTAGGGAACACAGACTAGTGAAGGTCT 540
481 CAATTCCGGAGTCTTGGGCGGAGAAACCAAGTAGGGAACACAGACTAGTGAAGGTCT 540
541 TCCCTGTGGGCAAGACACCCAGAGGACACACCCCTGACTTTCGGACATGCTTAAGT 600
541 TCCCTGTGGGCAAGACACCCAGAGGACACACCCCTGACTTTCGGACATGCTTAAGT 600
601 TTGAGAGTGACAAAGCACCAGGACATCTCTATGTGGAATATAGAGACACATTTCTCAACC 660
601 TTGAGAGTGACAAAGCACCAGGACATCTCTATGTGGAATATAGAGACACATTTCTCAACC 660
661 TGTCCCTGAAGAGTCTGTTTCTTAGTGAGGAGTCTCTGAGCACTTCCAGACAGAGT 720
661 TGTCCCTGAAGAGTCTGTTTCTTAGTGAGGAGTCTCTGAGCACTTCCAGACAGAGT 720
721 TTGCTCTCAAGGGCGATGATGAGTGTGTTGTAACACCCATCACATCTTAACTACTGA 780
721 TTGCTCTCAAGGGCGATGATGAGTGTGTTGTAACACCCATCACATCTTAACTACTGA 780
781 ATAGCTTATCCAAAGAGCAAGAGCTTGTTCATAGTGACGTGATCCACAATGCTG 840
781 ATAGCTTATCCAAAGAGCAAGAGCTTGTTCATAGTGACGTGATCCACAATGCTG 840
841 GGCCTCACGGGATTAAGAACTGAAGTACTACATCCCAAGAGTCTTACACCGGCTCT 900
841 GGCCTCACGGGATTAAGAACTGAAGTACTACATCCCAAGAGTCTTACACCGGCTCT 900
901 ACCACCGTATGCCGGGGTGGTGTCTCTACTCCGGCCCTTCCCTTGAGGCTGT 960
901 ACCACCGTATGCCGGGGTGGTGTCTCTACTCCGGCCCTTCCCTTGAGGCTGT 960
961 ACAGTCCGACTAGCCGGGTCCTCTCTACCTATTGATGATGTTTATACGGGAATGTGCC 1020
961 ACAGTCCGACTAGCCGGGTCCTCTCTACCTATTGATGATGTTTATACGGGAATGTGCC 1020
1021 TTCAGAACTGGGCTTGTTCAGAGAGCAACAAAGGCTTCAGACATTTGATTTGAAG 1080
1021 TTCAGAACTGGGCTTGTTCAGAGAGCAACAAAGGCTTCAGACATTTGATTTGAAG 1080
1081 AGAAAAATTAAGAAAAATATTGCTCTATATAGACTTAATGTTTACTATACAGCAAAAC 1140
1081 AGAAAAATTAAGAAAAATATTGCTCTATATAGACTTAATGTTTACTATACAGCAAAAC 1140
1141 CTCAGAGATGATTGATATCTGGTCTCAGTTCGAAAGTCTTAATTTAAATGCTGA 1196
1141 CTCAGAGATGATTGATATCTGGTCTCAGTTCGAAAGTCTTAATTTAAATGCTGA 1196
```

```
US-09-804-006-1
; Sequence 1, Application US/09804006
; Patent No. US20020119517A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jiahong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/126001
; CURRENT APPLICATION NUMBER: US/09/804,006
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/292,228
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1196
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-804-006-1
```

```
Query Match 100.0%; Score 1196; DB 10; Length 1196;
Best Local Similarity 100.0%; Pred No 0;
Matches 1196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AGATGAGTGTGGGCGTFCGAAGAGTCAAGTTGCTGGGCATCTCTGATGATGGCAATGTCT 60
Db 1 AGATGAGTGTGGGCGTFCGAAGAGTCAAGTTGCTGGGCATCTCTGATGATGGCAATGTCT 60
Qy 61 TCATTTATTTGATGTGGAGTCTCCAAAACAGTAGCCCAAGACAAATATGGAAGGGAG 120
Db 61 TCATTTATTTGATGTGGAGTCTCCAAAACAGTAGCCCAAGACAAATATGGAAGGGAG 120
Qy 121 GAGTAATAATCCGGAAGAGAGTTCTGGAAGCCACCAGCACATCCCGGGGCATACTGGA 180
Db 121 GAGTAATAATCCGGAAGAGAGTTCTGGAAGCCACCAGCACATCCCGGGGCATACTGGA 180
Qy 181 ACAGGGAACAGGAGAGCTGAACAGTGTGTACAAATCCCATCTTTGAACAGGTTGGCCAATC 240
Db 181 ACAGGGAACAGGAGAGCTGAACAGTGTGTACAAATCCCATCTTTGAACAGGTTGGCCAATC 240
Qy 241 AGACAGGGGAGCTAGCCACATCTCCAAACACAAAGTCACTGAGCTATTGTGAACAGACT 300
Db 241 AGACAGGGGAGCTAGCCACATCTCCAAACACAAAGTCACTGAGCTATTGTGAACAGACT 300
Qy 301 CGACGGTCATGACAGCTGTGACAGATTTTAATAATCTGCCGGACAGATTTTAAAGACTTTC 360
Db 301 CGACGGTCATGACAGCTGTGACAGATTTTAATAATCTGCCGGACAGATTTTAAAGACTTTC 360
Qy 361 TCTTGTATTGAGATGCCGGAATTTACTCGCTGCTTATAGATCAACCGAAGAAATGTGCAA 420
Db 361 TCTTGTATTGAGATGCCGGAATTTACTCGCTGCTTATAGATCAACCGAAGAAATGTGCAA 420
Qy 421 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAG 480
Db 421 AGAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTTCCAGAGGCAAG 480
Qy 481 CAATTCCGGAGTCTTGGGCGGAGAAACCAACCTAGGGAACACAGACTAGTGAAGGTCT 540
Db 481 CAATTCCGGAGTCTTGGGCGGAGAAACCAACCTAGGGAACACAGACTAGTGAAGGTCT 540
Qy 541 TCCTGTTGGGCAAGACACCCCGAGAGCAACCCCTGACCTTTCGGACATGCTTAAGT 600
Db 541 TCCTGTTGGGCAAGACACCCCGAGAGCAACCCCTGACCTTTCGGACATGCTTAAGT 600
Qy 601 TTGAGAGTGACAAAGCACCAGGACATCTCTGAGTGTGGAATATAGAGACACATTTCTCAACC 660
Db 601 TTGAGAGTGACAAAGCACCAGGACATCTCTGAGTGTGGAATATAGAGACACATTTCTCAACC 660
```



```
QY 123 GTAATAATCCCGAAGAGAGTCTTGGAAAGCACCAGCAGCTCCCGGGCATACTGGAAAC 182
Db 121 GTAATAATCCCGAAGAGAGTCTTGGAAAGCACCAGCAGCTCCCGGGCATACTGGAAAC 180
QY 183 AGGACAGAGAGAGTCTGACAGGTGGTACATCCCATCTTGAACAGGTGGCCAAATCAG 242
Db 181 AGGACAGAGAGAGTCTGACAGGTGGTACATCCCATCTTGAACAGGTGGCCAAATCAG 240
QY 243 ACAGGGAGCTAGCCACATCTCCAAACACCAAGTCACTGAGTATGTGAACAGCACTCG 302
Db 241 ACAGGGAGCTAGCCACATCTCCAAACACCAAGTCACTGAGTATGTGAACAGCACTCG 300
QY 303 ACGGTATGACAGCTGTGACAGATTTTAATATCTGCGGAGACAGATTTAAAGACTTTCTC 362
Db 301 ACGGTATGACAGCTGTGACAGATTTTAATATCTGCGGAGACAGATTTAAAGACTTTCTC 360
QY 363 TTGTATTTGAGATGCGGGAATTAATCTGCTGCTTATAGATCAACCGAAGAAATGTGCAAG 422
Db 361 TTGTATTTGAGATGCGGGAATTAATCTGCTGCTTATAGATCAACCGAAGAAATGTGCAAG 420
QY 423 AAGCCCTCTTACTATTGGCGATAAAGTCCCTCATTTCCACATTTTCCAGAAAGCAAGCA 480
Db 421 AAGCCCTCTTACTATTGGCGATAAAGTCCCTCATTTCCACATTTTCCAGAAAGCAAGCA 480
QY 483 ATTCGGAGCTCTGGGCGGAGAAACCAAGCTAGGAAACACAGCAGTAGTGGGTCTTC 542
Db 481 ATTCGGAGCTCTGGGCGGAGAAACCAAGCTAGGAAACACAGCAGTAGTGGGTCTTC 540
QY 543 CTGTTGGGCAAGACACCCCGAGGAGCAACACCTGACCTTTCGGACATGCTTAAGTTT 602
Db 541 CTGTTGGGCAAGACACCCCGAGGAGCAACACCTGACCTTTCGGACATGCTTAAGTTT 600
QY 603 GAGAGTGACAAAGCAGGACATCTCATGTGGAATATAGACACATTTCTCAACCTG 662
Db 601 GAGAGTGACAAAGCAGGACATCTCATGTGGAATATAGACACATTTCTCAACCTG 660
QY 663 TCCTTGAAGCAAGTGTGTTCTTAGTGGTGAGCAGCTTCTGTCACAGCAGAGTTT 722
Db 661 TCCTTGAAGCAAGTGTGTTCTTAGTGGTGAGCAGCTTCTGTCACAGCAGAGTTT 720
QY 723 GTCTTCAAGGGCGATGATGACGTGTTTGTGAACACCCATCAGATCCTTAATTAATGAT 782
Db 721 GTCTTCAAGGGCGATGATGACGTGTTTGTGAACACCCATCAGATCCTTAATTAATGAT 780
QY 783 AGCTTATCAAGAGCAAGCAAGCAAGTGTGTTATAGTGAAGTATCAGCAATGCTGG 842
Db 781 AGCTTATCAAGAGCAAGCAAGCAAGTGTGTTATAGTGAAGTATCAGCAATGCTGG 840
QY 843 CCTCAGCGGATAGAAACTGAAGTACTACATCCAGAAAGTCTTCTACACCGGCTCTAC 902
Db 841 CCTCAGCGGATAGAAACTGAAGTACTACATCCAGAAAGTCTTCTACACCGGCTCTAC 900
QY 903 CCACCGTATCCGGGGGTGGTGAATTCCTGTAAGTCCCGGCGCCCTTGAGGCTGTAC 962
Db 901 CCACCGTATCCGGGGGTGGTGAATTCCTGTAAGTCCCGGCGCCCTTGAGGCTGTAC 960
QY 963 AGTGCAGACTAGCCGGGTGCTCATCTACCTATGATGATGTTTATACGGGAATGCGCTT 1022
Db 961 AGTGCAGACTAGCCGGGTGCTCATCTACCTATGATGATGTTTATACGGGAATGCGCTT 1020
QY 1023 CAGAAACTGGGCTTCTTCAGAGAGACCAAGGCTTCAGGACATTTGATATTGAAGAG 1082
Db 1021 CAGAAACTGGGCTTCTTCAGAGAGACCAAGGCTTCAGGACATTTGATATTGAAGAG 1080
QY 1083 AAAAATAAGAAAAATATTTGTTCTTATAGACCTAAAGTGTAGTACATAGCAGAAAACT 1142
Db 1081 AAAAATAAGAAAAATATTTGTTCTTATAGACCTAAAGTGTAGTACATAGCAGAAAACT 1140
QY 1143 CAAGAGATGATGATATCTGGTCTCAGTTGCAAAAGTCTCAATTTAAATATGC 1193
Db 1141 CAAGAGATGATGATATCTGGTCTCAGTTGCAAAAGTCTCAATTTAAATATGC 1191
```

```
RESULT 6
US-09-804-357-13
; Sequence 13, Application US/09804357
; Patent No. US2001002480A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/109001
; CURRENT APPLICATION NUMBER: US/09/804,357
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/195,896
; PRIOR FILING DATE: 1998-11-19
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 1707
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (246)...(1436)
; NAME/KEY: misc feature
; LOCATION: (1)...(1707)
; OTHER INFORMATION: n = A,T,C or G
US-09-804-357-13
```

```
Query Match 75.8%; Score 906.4; DB 9; Length 1707;
Best Local Similarity 84.9%; Pred. No. 4.3e-275;
Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 1 AGATGAGTGTGGGCGTGAAGAGTCAAGTGTGCTGGGCATCCTGATGCGCAATGTCT 60
Db 244 AAATGAGTGTGGGCGTGAAGAGTAAAGTGTGTTGGGTATCTCTGATGCGCAATGTCT 303
QY 61 TCATTATTGATGTGGAAGTCTCCAAAGACAGTAGCAAGACAAATAAGTGAAGGAG 120
Db 304 TCATTATTATTATGGAAGTCTCCAAAGACAGTAGCAAGACAAATAAGTGAAGGAG 363
QY 121 GAGTAATAATCCGAAGAGAGTGTGGAAGACCCAGCAGCTCCCGGGCATACTGGA 180
Db 364 AAGTAATAATCCGAAGAGAGTGTGGAAGATATCTACCCCTCCCGAGGCATACTGGA 423
QY 181 ACAGGGAACAGAGAGAGTGAACAGGTGCTACATCCCATCTTGAACAGGTGGCCATC 240
Db 424 ACCGAGACAGAGAGAGTGAACAGGTGCTACACCCCATCTCTGAGCATGCTGACCAAC 483
QY 241 AGACAGGGAGTAGCCACATCTCCAAACACAAAGTCACTGAGCTATTGTGAACAGACT 300
Db 484 AGACAGGGAGGGCGGCGAGGCTCTCCATATATAGCCATCTGAACTACTGCAACCTGAC 543
QY 301 CGACGGTCAATGACAGCTGTGACAGATTTTAATAATCTGCGGAGCAGATTTAAAGACTTC 360
Db 544 TGAGGTCACGTCGGTGGTGTACGGGTTTAAACAACCTTCCCGGAGAGATTTAAAGACTTC 603
QY 361 TCTTGTATTGAGATGCGGGAATTAATCTGCTGCTTATAGATCAACCGAAGAAATGTGCA 420
Db 604 TGCTGTATTGAGATGCGGCAATTAATCACTGCTTATAGATCAACCGGATTAAGTGTGCA 663
QY 421 AGAAGCCCTCTTACTATTGGGATAAAGTCCCTCATTTCCACATTTTCCCAAGAGCAAG 480
Db 664 AGAAGCCCTTCTTGTGCTGGGATTAAGTCCCTCATTTCCACATTTTCCCAAGAGCAAG 723
QY 481 CAATTCGGGAGTCTTGGGGCGGAGAAACCAAGTAGGGAACCAAGCAGTACTAGGAGTCT 540
Db 724 CAATCCGGGAATCTTGGGGCGGAGAAAGCAAGCAGGAGGAACCAAGAGTGTGCGAGTCT 783
```



```
Db 1204 ACATATCACTGACAGGTCATCTCTACCCCATGATGACGCTTTATAGTGAATGTGCC 1263
QY 1021 TTCAAGAACTGGGCTGTGTCCAGAGAGCAACAAAGGCTTCAGGACATTTGATGAG 1080
Db 1264 TTCAAGAACTGGGCTGTGTCCAGAGAGCAACAAAGGCTTCAGGACATTTGATGAG 1323
QY 1081 AGAAATTAAGAAATATTTGTTCTATATAGACCTAATGTTAGTACATGACAAAC 1140
Db 1324 AGAAATTAAGAAATATTTGTTCTATATAGACCTAATGTTAGTACATGACAAAC 1383
QY 1141 CTCAGAGATGATTGATATCTGTCTCAGTTCGAAGTCTTAATTTAAATGCTGA 1196
Db 1384 CTCAGAGATGATTGATATTTGTTCTCAGTTCGAAGTCTTAATTTAAATGCTGA 1439

RESULT 8
US-10-109-563-3
; Sequence 3, Application US/10109563
; Publication No. US20020170075A1
; GENERAL INFORMATION:
; APPLICANT: Leviten, Michael W.
; TITLE OF INVENTION: TRANSGENIC MICE COMPRISING
; FILE REFERENCE: R-021
; CURRENT APPLICATION NUMBER: US/10/109,563
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: US 60/280,706
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1831
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-109-563-3

Query Match 75.8%; Score 906.4; DB 13; Length 1831;
Best Local Similarity 84.9%; Pred. No. 4.5e-275;
Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 1 AGATGAGTGTGGGCTCCAGAGTCAAGTCTGCTGGGCTCTCTGATGATGGCAATGTCT 60
Db 234 AATGAGTGTGGGCTCCAGAGTCAAGTCTGCTGGGCTCTCTGATGATGGGCAATGTCT 293
QY 61 TCATTATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 120
Db 294 TCATTATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 353
QY 121 GAGTAATATCCGGAAGAGAGTCTGGAAGCCACCCAGCACTCCCGGGCATACTGGA 180
Db 354 AGTAATATATCCGGAAGAGAGTCTGGAAGCCACCCAGCACTCCCGGGCATACTGGA 413
QY 181 ACAGGGAACAGGAGAGTCTGGAAGCCACCCAGCACTCCCGGGCATACTGGA 240
Db 414 ACCGAGAGAGAGAGTCTGGAAGCCACCCAGCACTCCCGGGCATACTGGA 473
QY 241 AGACAGGAGAGTCTGGAAGCCACCCAGCACTCCCGGGCATACTGGA 300
Db 474 AGACAGGAGAGTCTGGAAGCCACCCAGCACTCCCGGGCATACTGGA 533
QY 301 CGAGGTCATGAGAGTCTGGAAGCCACCCAGCACTCCCGGGCATACTGGA 360
Db 534 TGAGGTCATGAGAGTCTGGAAGCCACCCAGCACTCCCGGGCATACTGGA 593
QY 361 TCTGTTGTTGAGATCCGGAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 420
Db 594 TGTGTTGTTGAGATCCGGAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 653
QY 421 AGAAGCCCTTCTTACTATTGGGCAATAAGTCCCTCAATTCACATTTTGGCAGAGCAAG 480
Db 654 AGAAGCCCTTCTTACTATTGGGCAATAAGTCCCTCAATTCACATTTTGGCAGAGCAAG 713
```

```
QY 481 CAATTCGGGAGTCTTGGGCGGAGAAACCAACAGTAGGGAACCAACAGTACTGAGGGTCT 540
Db 714 CAATTCGGGAGTCTTGGGCGGAGAAACCAACAGTAGGGAACCAACAGTACTGAGGGTCT 773
QY 541 TCCTGTTGGGCAAGACACCCCGAGAGCAACCAACAGTACTGAGGGTCTTGGGAGTCTTAAGI 600
Db 774 TCCTGTTGGGCAAGACACCCCGAGAGCAACCAACAGTACTGAGGGTCTTGGGAGTCTTAAGI 833
QY 601 TTGAGAGTGACAAGCAACCCAGGACATCTCTCATGTGGAACATATAGAGACATCTTCAACC 660
Db 834 TTGAGAGTGACAAGCAACCCAGGACATCTCTCATGTGGAACATATAGAGACATCTTCAACC 893
QY 661 TGTCCCTGAAGAGTGTCTGTTTCTTAGGTGGTGGGAGTCTCTGTCAGGAGTCTGTCAGGAGT 720
Db 894 TGTCTGAAGAGTGTCTGTTTCTTAGGTGGTGGGAGTCTCTGTCAGGAGTCTGTCAGGAGT 953
QY 721 TTGCTTCAAGGGGATGATGACGTGTTTGTGAACCCATCAGATCTCTTAATTAATCTGA 780
Db 954 TTGCTTCAAGGGGATGATGACGTGTTTGTGAACCCATCAGATCTCTTAATTAATCTGA 1013
QY 781 ATAGCTTATCCAAGAGCAACCAAGCAAGTCTCTCATAGGTGATGATCCACAATGCTG 840
Db 1014 ATAGCTTATCCAAGAGCAACCAAGCAAGTCTCTCATAGGTGATGATCCACAATGCTG 1073
QY 841 GGCCTCACCGGATGAAGACTGAAGTACTACATCCCAAGAGTCTCTTACACCGGCTCT 900
Db 1074 GACCTCATCGGGATGAAGACTGAAGTACTACATCCCAAGAGTCTCTTACACCGGCTCT 1133
QY 901 ACCCACTGATGCGGGGGTGTGATTCCTGATCCGCGCCCTTCCCTTGGAGTGT 960
Db 1134 ACCCACTGATGCGGGGGTGTGATTCCTGATCCGCGCCCTTCCCTTGGAGTGT 1193
QY 961 ACAGTGCAGTACGGGCTGCTCTCTACCTTATGATGATGTTTATACGGGAATGTGCC 1020
Db 1194 ACCATATCACTGACAGGTCATCTCTACCCATGATGACGCTTTATAGTGAATGTGCC 1253
QY 1021 TTCAGAACTGGGCTTGTTCAGAGAGCAACAAAGGCTTCAGGACATTTGATATGAG 1080
Db 1254 TTCAGAACTGGGCTTGTTCAGAGAGCAACAAAGGCTTCAGGACATTTGATATGAG 1313
QY 1081 AGAAATTAAGAAATATTTGTTCTATATAGACTAATGTTAGTACATAGCAGAAAC 1140
Db 1314 AGAAATTAAGAAATATTTGTTCTATATAGACTAATGTTAGTACATAGTAAAC 1373
QY 1141 CTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1196
Db 1374 CTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1429
```

RESULT 9

```
US-09-972-912-1
; Sequence 1, Application US/09972912
; Patent No. US20020110867A1
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R.
; RUBEN, STEVEN M.
; TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/972,912
```

```
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/049,022
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0620001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2745 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 233..1423
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 233..328
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 329..1423
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-972-912-1

Query Match
Best Local Similarity 84.9%; Score 906.4; DB 10; Length 2745;
Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 1 AGATGAGTGTGGGCGCTCGAAGAGTCAGTTCCTGGGCATCCGTGATGATGGCAAAATGTCT 60
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 231 AAATGAGTGTGGACGCTCGAAGAAATAAGTTTGTGGTATCTCGTATGATGGCAAAATGTCT 290
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 61 TCATTATTATTGTTGGAGTCTCCAAAACAGTAGCCCAAGACAAAATAATGAAAGGGAG 120
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 291 TCATTATTATTGTTGGAGTCTCCAAAACAGTAGCCCAAGACAAAATAATGAAAGGGAG 350
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 121 GAGTAATAATCCGGAAGAGAGTCTGGAAGCCACCCAGCACTCCCGGGGCATCTACTGGA 180
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 351 AAGTAATAATACCAAGAGAGAGTCTGGAAGATATCTACCCCTCCCGAGGCATCTACTGGA 410
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 181 ACAGGGAACAGAGAGAGTCTGAGAGTGTGACAAATCCCATCTGACAGAGTGGCCAACT 240
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 411 ACCGAGAGCAAGAGAGTCTGAGAGTGTGACAAATCCCATCTGAGAGTGGCCAACT 470
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 241 AGACAGGGAGTCTGAGAGTCTGAGAGTGTGACAAATCCCATCTGAGAGTGGCCAACT 300
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 471 AGACAGGGAGGGGGGCGAGGCTCTCCCAATATAGGCACTGAGAGTGGCCAACTGACC 530
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 301 CGAGGTCATGAGAGTGTGAGAGTGTGACAAATCCCATCTGAGAGTGGCCAACTTTC 360
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 531 TGAGGGTCTGAGTGGTGTGAGAGTGTGACAAATCCCATCTGAGAGTGGCCAACTTTC 590
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 361 TCTTGATTGAGATCGGGAATCTGCTCTATAGATCAACCGAAGAAATGTGCA 420
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 591 TCGTATTGAGATCGGGAATCTGCTCTATAGATCAACCGAAGAAATGTGCA 650
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 421 AGAAGCCCTCTTACTATTGGGATTAAGTCCCTCATCTGAGAGTGGCCAACTGAGAG 480
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 651 AGAAACCTCTTCTGCTGGGATTAAGTCCCTCATCTGAGAGTGGCCAACTGAGAG 710
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 481 CAATTCGGAGTCTTGGGCGGAGAAACCAACCTAGGAGAACAGACAGATGAGAGGTCT 540
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 711 CAATTCGGGAATCTGCGGCAAGAAAGCAACGAGGAAACCAACGAGTGGTGGAGTCT 770
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 541 TCCGTGTGGCAAGACACCCCGAGAGCAACCCCTGACCTTCGGACATGCTTAAGT 600
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

```
DB 771 TCCTGTGGCCAGACACACCCCGAGAGCAACACCCCGACCTTTTCAGATATGCTCAAT 830
QY 601 TTGAGAGTGAAGACACCGAGGACATCTCATGTGGAACTATAGAGACACATTTCTCAACC 660
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 831 TTGAGAGTGAAGACACCGAGGACATCTCATGTGGAACTATAGAGACACATTTCTCAACT 890
QY 661 TGTCCCTGAAGGAAGTGTCTTTTCTAGGTGGGTGAGCAGCTTCCCTGAGAGGAGAGT 720
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 891 TGTCTCTGAAGGAAGTGTCTTTCTCAGGTGGGTAACTTCTCTCCGCCAGACATGAT 950
QY 721 TTGCTTTCAAGGGCGATGATGAGCTGTTTGTGAACACCCATCACATCTTAATCTTGA 780
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 951 TTGTTTCAAGGGCGATGAGCTGTTTCTGGAACCCATCACATCTTGAATCTTGA 1010
QY 781 ATAGCTTATCAAGAGCAAGCAAGACTTGTTCATAGTGTGACGTGATCCCAATGCTG 840
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 1011 ATAGTTTATCAAGACCAAGCAAGATCTCTCATAGTGTGATGTGATCCCAATGCTG 1070
QY 841 GGCTCACCGGGATAAGAACTGAACTACTACATCCCAAGAGTCTTCTACACGGGCTCT 900
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 1071 GACCTCATCGGATAGAGCTGAACTACTACATCCCAAGAGTGTTCCTCTGAGCTCT 1130
QY 901 ACCACCGTATCGGGGGTGGTTCCTGATTCCTGCTGACCTGATGATGATGATGATGATG 960
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 1131 ACCACCGTATCGGGGGTGGTTCCTGATTCCTGCTGACCTGATGATGATGATGATGATG 1190
QY 961 ACAGTGGCTAGCGGGTCCATCTCTACCTATTGATGATGATGATGATGATGATGATGATG 1020
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 1191 ACCATATCTGACCGAGGTCCTCTCTACCTATTGATGATGATGATGATGATGATGATGATG 1250
QY 1021 TTCAGAACTGGGCTTTTCCAGAGCAAGCAAGAGCTTCAGACATTTGATATTTGAAG 1080
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 1251 TTCAGAACTGGGCTTTTCCAGAGCAAGCAAGAGCTTCAGACATTTGATATTTGAAG 1310
QY 1081 AGAAAAATAGAAAAATATTTGTTCTATATAGACCTATGTTAGTACATAGCAGAAAC 1140
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 1311 AGAAAAATAGAAAAATATTTGTTCTATATAGACCTATGTTAGTACATAGCAGAAAC 1370
QY 1141 CTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1196
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 1371 CTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1426

RESULT 10
US-09-972-912-20
; Sequence 20, Application US/09972912
; Patent No. US20020110867A1
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R.
; RUBEN, STEVEN M.
; TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/972,912
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/049,022
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
```

```
;
;
;   REGISTRATION NUMBER: 36,688
;   REFERENCE/DOCKET NUMBER: 1488.0620001
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 371-2600
;   TELEFAX: (202) 371-2540
;   INFORMATION FOR SEQ ID NO: 20:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 857 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: both
;   TOPOLOGY: both
;   MOLECULE TYPE: cDNA
;   SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-972-912-20

Query Match      24.1%; Score 288.8; DB 10; Length 857;
Best Local Similarity 85.0%; Pred. No. 3.6e-80;
Matches 345; Conservative 0; Mismatches 59; Indels 2; Gaps 2;

QY 787 TATCCAGAGCAAGCCAAAGACTGTTTCATAGGTGACGTGATCCACAATGCTGGGCGCTC 846
Db 20 TATTCAGAGCAAGCCAAAGACTGTTTCATAGGTGATGATCCACAATGCTGGACCTC 79
QY 847 ACCGGGATAGAACTGAAGTACTACATCCCAAGTCTTCTACACCGGCGTCTACCCAC 906
Db 80 ATCGGGATAGAAAGCTGAAGTACTACATCCCAAGTCTTCTACTCJGGCCTCTACCCAC 139
QY 907 CGTATCGCGGGGTGGTGGATTCCTGTACTCGGGCCCTTGGCCTTGAGGCTGTACAGTG 966
Db 140 CCTATCGAGGGGAGGGGGGTTCTCTACTCCGGCCACCTGGGCCCTGAGGCTGTACCAT 199
QY 967 CGACTAGCGGGTCCATCTCTACCTCTATGATGATGATGATGATGATGATGATGATGAT 1026
Db 200 TCACTGACAGGTCCTCTCTACCTCTATGATGATGATGATGATGATGATGATGATGATG 259
QY 1027 AACTGGGCGCTTGTCCAGAGGACCAAGGCTTCAGGACATTTGATATTGAAGAGAAAA 1086
Db 260 AACTCGGCGCTCGTCCAGAGAACACAAAGGCTTCAGGACATTTGATATCGAGGAGAAA 319
QY 1087 ATAGAGAAATATTTGCTCTATAGACCTATGTTAGTACATAGA-GAANAACCTCAA 1145
Db 320 AAAAAATACATCTCTCTCTATGATGATGATGATGATGATGATGATGATGATGATGATG 379
QY 1146 GAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1190
Db 380 GAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 425

RESULT 11
US-09-972-912-26
; Sequence 26, Application US/09972912
; Patent No. US20020110867A1
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R.
; RUBEN, STEVEN M.
; TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version: #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/972.912
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
```

```
;
;
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 09/049,022
;   FILING DATE: <Unknown>
;   ATTORNEY/AGENT INFORMATION:
;   NAME: STEFFE, ERIC K.
;   REGISTRATION NUMBER: 36,688
;   REFERENCE/DOCKET NUMBER: 1488.0620001
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 371-2600
;   TELEFAX: (202) 371-2540
;   INFORMATION FOR SEQ ID NO: 26:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 480 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: both
;   TOPOLOGY: both
;   MOLECULE TYPE: cDNA
;   SEQUENCE DESCRIPTION: SEQ ID NO: 26:
US-09-972-912-26

Query Match      19.7%; Score 236; DB 10; Length 480;
Best Local Similarity 76.2%; Pred. No. 1.2e-63;
Matches 343; Conservative 0; Mismatches 97; Indels 10; Gaps 4;

QY 302 GACGGTCATGACAGCTGTGACAGATTTTAATATCTGCCGGACAGATTTAAGACTTTCT 361
Db 2 GAGGTCACGTCGGTGGTTACGGGTTTTAACAACTTGGCGGACAGATTTAAAGACITTTCT 61
QY 362 CTTGTATTGAGATGCCGAATTACTCGCTTATAGATCAACCGAAGAAATGTGCAAA 421
Db 62 GCTGTATTGAGATGCCGAATTACTCGCTTATAGATCAACCGGATAGTGTGCAAA 121
QY 422 GAAGCCCTTCTTACTATTGGCGATAAAGTCCCTCATTCACATTTTGCAGAGGCAA-G 480
Db 122 GAAACCTTTTCTTGTGCTGGCGATTAAAGTCCCTCACTCCACATTTTGCAGAGGCAAG 181
QY 481 CAATTCGGGAGTCTGGGGCCGAGAACCAACGTTAGGAAACCAAGACAGACTAGTGGGTCT 540
Db 182 CAATTCGGGAAATCTTGGGGCCAGAAAGCAACGAGGAAACCAACGGTGGTGGCGAGTCT 241
QY 541 TCCTGTGGGCAAGACACCCCGAGAGCAACCCCTGACCTTTTCGGACATGCT-TAAG 599
Db 242 TCCTGCTGGCGAGACACCCCGAGAGCAACCCCGACCTTTTCAGATATGCTGAAT 301
QY 600 TTTGAGAGTGAACGACCCAGGACATCTCATGTGG-----AACTATAGAGACACATCTT 655
Db 302 TTTGAGAGTGAAGACCAACCAAGACATTTCTTATGTGGGAACCTACAGAGGACACTTTCTTC 361
QY 656 CAACCTGTCCCTGAAGGAAGTGTCTTCTTAGTGGGTG-----AGCACTTCTCTCCAG 711
Db 362 AANTTGTCTNTGGAAGGAAGTGTCTTCTTTTCAGGTGGGTTAAGTTATTTCTCTGCCAG 421
QY 712 ACGCAGAGTTTGTCTTCAAGGGCGAIGATG 741
Db 422 ACATTGAGTTTGTCTTTTCAAGGGCGGATG 451

RESULT 12
US-09-972-912-11
; Sequence 11, Application US/09972912
; Patent No. US20020110867A1
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R.
; RUBEN, STEVEN M.
; TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/09/972,912
FILING DATE: 10-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/049,022
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
FILING DATE: <Unknown>
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0620001
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-972-912-11

Query Match 18.8%; Score 224.6; DB 10; Length 282;
Best Local Similarity 86.9%; Pred. No. 3.3e-60;
Matches 245; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 464 TTTTCCAGAGGCAAGCAATCGGAGTCTTGGGCGGAGAACCAACGTAAGGAACCA 523
Db 1 TTTTCCAGAGGCAAGCAATCGGAGTCTTGGGCGGAGAACCAACGTAAGGAACCA 60

Qy 524 GACAGTAGTGAGGCTTCTCTGTTGGGCAAGACACCCCGAGGACATCTCATGTGGAACATAG 583
Db 61 AACGGTGTGGAGTNTTCTGTTGGGCGGAGAACCAACGTAAGGAACCA 120

Qy 584 TTCCGACATCTTAAGTTTGAGAGTGACAGACACCCCGAGGACATCTCATGTGGAACATAG 643
Db 121 TTCAGATATCTGAAATTTGAGAGTGAGAGACACCCCGAGGACATCTCATGTGGAACATAG 180

Qy 644 AGACACATCTTCAACCTGCTCCCTGAGAGAGTGTCTTCTAGGTGGGTGAGCACTTC 703
Db 181 AGACACATCTTCAACCTGCTCTGAGGAGAGTGTCTTCTAGGTGGGTGAGCACTTC 240

Qy 704 CTGTCCAGACGACAGTTTGTCTTCAAGGCGGATGACGT 745
Db 241 CTGCCCAGACACTGAGTTTGTCTTCAAGGCGGATGACGT 282

RESULT 13
US-09-972-912-34
Sequence 34, Application US/09972912
Patent No. US20020110867A1
GENERAL INFORMATION:
APPLICANT: SOPPET, DANIEL R.
RUBEN, STEVEN M.
TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
COMPUTER: IBM PC compatible
CURRENT APPLICATION DATA:

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/972,912
FILING DATE: 10-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/049,022
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
FILING DATE: <Unknown>
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0620001
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-09-972-912-34

Query Match 18.8%; Score 224.6; DB 10; Length 282;
Best Local Similarity 86.9%; Pred. No. 3.3e-60;
Matches 245; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 464 TTTTCCAGAGGCAAGCAATCGGAGTCTTGGGCGGAGAACCAACGTAAGGAACCA 523
Db 1 TTTTCCAGAGGCAAGCAATCGGAGTCTTGGGCGGAGAACCAACGTAAGGAACCA 60

Qy 524 GACAGTAGTGAGGCTTCTCTGTTGGGCAAGACACCCCGAGGACATCTCATGTGGAACATAG 583
Db 61 AACGGTGTGGAGTNTTCTGTTGGGCGGAGAACCAACGTAAGGAACCA 120

Qy 584 TTCCGACATCTTAAGTTTGAGAGTGACAGACACCCCGAGGACATCTCATGTGGAACATAG 643
Db 121 TTCAGATATCTGAAATTTGAGAGTGAGAGACACCCCGAGGACATCTCATGTGGAACATAG 180

Qy 644 AGACACATCTTCAACCTGCTCCCTGAGAGAGTGTCTTCTAGGTGGGTGAGCACTTC 703
Db 181 AGACACATCTTCAACCTGCTCTGAGGAGAGTGTCTTCTAGGTGGGTGAGCACTTC 240

Qy 704 CTGTCCAGACGACAGTTTGTCTTCAAGGCGGATGACGT 745
Db 241 CTGCCCAGACACTGAGTTTGTCTTCAAGGCGGATGACGT 282

RESULT 14
US-09-972-912-31
Sequence 31, Application US/09972912
Patent No. US20020110867A1
GENERAL INFORMATION:
APPLICANT: SOPPET, DANIEL R.
RUBEN, STEVEN M.
TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/972,912
FILING DATE: 10-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/049,022
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0620001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-09-972-912-31
Query Match 18.7%; Score 223.6; DB 10; Length 303;
Best Local Similarity 86.4%; Pred. No. 7.2e-60;
Matches 247; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 342 GACAGATTTAAAGACITTCCTCTGTTATTTGAGATGCGGGAATTAAGTCTGCTTATAGAT 401
Db 1 GACAGATTTAAAGACITTCCTCTGTTATTTGAGATGCGGGAATTAAGTCTGCTTATAGAT 60
QY 402 CAACCGAAGAAATGTGCAAGAAGCCTCTCTTACTATTTGGCGATGAAGTCCCTCATTTCCA 461
Db 61 CAGCCGGATAGTGTGCAAGAAGCCTCTCTTACTATTTGGCGATGAAGTCCCTCATTTCCA 120
QY 462 CATTTTCCCAAGGCAAGCAATTCGGGAGTCTTGGGGCCGAGAAACCAACGTAAGGGAAC 521
Db 121 CATTTTCCCAAGGCAAGCAATTCGGGAGTCTTGGGGCCGAGAAACCAACGTAAGGGAAC 180
QY 522 CAGACAGTAGTAGGCTTCTCTGTTGGGCAAGACACCCCGAGGACCAACGACCCCTGAC 581
Db 181 CAACCGTGTGGGAGTCTCTCTGTTGGGCAAGACACCCCGAGGACCAACGACCCCGAC 240
QY 582 CTTTCGACATGCTTAAAGTTTGGAGTGACAAAGCAGCAGGACATCC 627
Db 241 CTTTCAGATATGCTGAATTTGAGAGTTAGAGCACCACCAAGACATTC 286

RESULT 15
US-09-972-912-12
Sequence 12, Application US/09972912
Patent No. US20020110867A1
GENERAL INFORMATION:
APPLICANT: SOPPET, DANIEL R.
RUBEN, STEVEN M.
TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.D.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/972,912
FILING DATE: 10-Oct-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/049,022
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0620001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 266 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-972-912-12
Query Match 17.7%; Score 211.2; DB 10; Length 266;
Best Local Similarity 86.9%; Pred. No. 5.5e-56;
Matches 231; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 464 TTTTGCAGAGGCAAGCAATTCGGGAGTCTTGGGGCCGAGAAACCAACGTAGGGAACCA 523
Db 1 TTTTGCAGAGGCAAGCAATTCGGGAGTCTTGGGGCCGAGAAACCAACGTAGGGAACCA 60
QY 524 GACAGTAGTAGGCTTCTCTGTTGGGCAAGACACCCCGAGGACCAACCCCTGACCT 583
Db 61 AACGGTGTGGGAGTNTCTCTGCTGGGCGAGACACCCCGAGGACCAACCCCGACCT 120
QY 584 TTCGGACATGCTTAAGTTTGAGAGTGACAAAGCAGGACATCCTCATGTGGAAGTATAG 643
Db 121 TTCAGATATGCTGAATTTGAGAGTNAGAAGCACCACCAAGACATTTCTATGTGAAGTACAG 180
QY 644 AGACACATTTCTCAACCTGTCCTGAAGGAAGTGTCTTCTTAGTGGTGGTGAGCACTTC 703
Db 181 AGACATTTTCTCAACTTGTCTCTGAGGAAGTGTCTTCTCAGTGAGTGTGTAAGTACTTC 240
QY 704 CTGTCCAGACGACAGAGTTTGTCTTCA 729
Db 241 CTGCCCAGACACTGAGTTTGTCTTCA 266
Search completed: October 20, 2003, 22:44:22
Job time : 334.002 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2003, 13:13:22 ; Search time 6381.7 Seconds
(without alignments)
10942.656 Million cell updates/sec

Title: US-09-804-357B-13
Perfect score: 1707
Sequence: 1 acgcgtccgcgcagcgcag.....aatctgttggatggccctt 1707

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl :

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_em.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rnd.*
- 36: em_htg_man.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1534.8	89.9	2811	9	BC047933	Homo sapi
2	1531.4	89.7	1831	9	AF092051	Homo sapi
3	1509.2	88.4	2714	9	AF288208	Homo sapi
4	1504.2	88.1	1912	6	BD093191	Useful po
5	1504.2	88.1	1912	9	AB049584	Homo sapi
6	1311.4	76.8	95088	9	AC093401	Homo sapi
7	1309.4	76.7	2454	9	BC030579	Homo sapi
8	1269.8	74.4	2516	9	AF288209	Homo sapi
9	1231.4	72.1	2442	6	BD156964	Primer fo
10	1231.4	72.1	2442	9	AK002009	Homo sapi
11	1181	69.2	1197	9	HSA6077	Homo sapi
12	1002.2	58.7	2672	10	AY043479	Mus muscu
13	954.2	55.9	200191	9	AL390027	Human DNA
14	954.2	55.9	288669	2	AL390027	Homo sapi
15	941.2	55.1	2468	10	BC009075	Mus muscu
16	940.2	55.1	210863	10	AL772364	Mouse DNA
17	931.4	54.6	1260	10	AF092050	Mus muscu
18	925.4	54.2	178009	2	AC119549	Rattus no
19	917.4	53.7	318862	2	AC109547	Rattus no
20	910.6	53.3	172509	2	AC024716	Homo sapi
21	717.2	42.0	832	6	BD149792	Primer fo
22	577.6	33.8	126480	2	AC040954	Mus muscu
23	533.6	31.3	556	6	BD071997	Secreted
24	425.2	24.9	170370	2	BX469915	Danio rer
25	423.2	24.8	2162	5	AF321831	Danio rer
26	414.2	24.3	197598	2	AC144887	Bos tauru
27	379	22.2	1682	5	AF321828	Danio rer
28	229.8	13.5	189326	9	AC018462	Homo sapi
29	223.4	13.1	473	6	BD052134	Sequence
30	203.4	11.9	1245	6	AX513603	Sequence
31	192.6	11.3	1564	5	AF321829	Danio rer
32	176.8	10.4	230799	2	AC097668	Rattus no
33	176.4	10.3	2260	5	AF321827	Danio rer
34	174.4	10.2	1134	6	AR202302	Sequence
35	174.2	10.2	209192	2	AC102609	Mus muscu
36	171	10.0	1502	10	AF502429	Mus muscu
37	171	10.0	2367	10	BC031187	Mus muscu
38	167.8	9.8	1360	6	AX451423	Sequence
39	166.2	9.7	1434	9	AF502430	Homo sapi
40	166.2	9.7	168880	9	AC017104	Homo sapi
41	164.8	9.7	2003	6	AX513601	Sequence
42	163.2	9.6	1116	6	BD063610	Human pro
43	163.2	9.6	1119	9	HSA130847	Homo sapi
44	163.2	9.6	1144	9	AF293973	Homo sapi
45	163.2	9.6	2186	6	BD063635	Human pro

ALIGNMENTS

RESULT 1
BC047933
LOCUS
DEFINITION
BC047933 Homo sapiens UDP-GlcNAc:betaGal
beta-1,3-N-acetylglucosaminyltransferase 1, transcript variant 1,
mRNA (cDNA clone MGC:49838 IMAGE:5767816), complete cds.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
BC047933
MGC.
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2811)
AUTHORS
Straussberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,

Db	842	AGTGAGAGACCAACAGACATCTTATGTGAACTACAGAGACACTTCTCAACTTGCT	901
QY	909	CTGAAGGAAGTGTGTTCTCAGTGGGTAAAGTACTTCCIGCCGACACTGAGTTTGT	968
Db	902	CTGAAGGAAGTGTGTTCTCAGTGGGTAAAGTACTTCCIGCCGACACTGAGTTTGT	961
QY	969	TTCAAGGGCGATGACCATGTTTGTGACACCCATCACAICTGTAATCTGTAATG	1028
Db	962	TTCAAGGGCGATGACCATGTTTGTGACACCCATCACAICTGTAATCTGTAATG	1021
QY	1029	TTATCAAGACCAACCAAGATCTCTCATAGGTGATGTGATCCACAACTCTGACCT	1088
Db	1022	TTATCAAGACCAACCAAGATCTCTCATAGGTGATGTGATCCACAACTCTGACCT	1081
QY	1089	CATCGGGATGAAGGTGAAGTACTACATCCCAAGAGTTGTTTACCTCTGACCTAC	1148
Db	1082	CATCGGGATGAAGGTGAAGTACTACATCCCAAGAGTTGTTTACCTCTGACCTAC	1141
QY	1149	CCCTATGCGAGGGAGGGGTTCCCTACTCCGGCCACTGGCCCTGAGGCTGACCAT	1205
Db	1142	CCCTATGCGAGGGAGGGGTTCCCTACTCCGGCCACTGGCCCTGAGGCTGACCAT	1201
QY	1209	ATCACTGACGAGTCCATCTCTACCCCATTTGATGAGCTTTATACCTGGAATGTCCT	1265
Db	1202	ATCACTGACGAGTCCATCTCTACCCCATTTGATGAGCTTTATACCTGGAATGTCCT	1261
QY	1269	AACTCGGCTCTGTCAGAGAACACAAAGGCTTCAGACATTTGATATCGAGGAGAA	1328
Db	1262	AACTCGGCTCTGTCAGAGAACACAAAGGCTTCAGACATTTGATATCGAGGAGAA	1321
QY	1329	ACAAAAATACATCTGCTCTATGATGATCTGATCTAGTATGATAGTAAAGCTCAA	1388
Db	1322	ACAAAAATACATCTGCTCTATGATGATCTGATCTAGTATGATAGTAAAGCTCAA	1381
QY	1389	GAGATGATTGATATTTGGTCTCAGTTCAGAGTGTCTATTTAAATGCTAAATAGAT	1448
Db	1382	GAGATGATTGATATTTGGTCTCAGTTCAGAGTGTCTATTTAAATGCTAAATAGAT	1441
QY	1449	AACTCAATTTKSGATGGAAGGGTWTGTTGATGAGTGGCCATGTTGGGCTCATAT	1508
Db	1442	AACTCAATTTKSGATGGAAGGGTWTGTTGATGAGTGGCCATGTTGGGCTCATAT	1501
QY	1509	AGAGTAATTTCTATTNAAC--ATGAAATGTCCTTATGATGATACCATTT--ANGC	1565
Db	1502	AGAGTAATTTCTATTNAAC--ATGAAATGTCCTTATGATGATACCATTT--ANGC	1561
QY	1566	CTTAANCCCTT---CATTTGACTACGTGAAGGAGGAAAGCGGAGAGGTAATTTT	1622
Db	1562	CTTAANCCCTTCAATTTGGTACTACGTGAAG-AGGGAAGCGGAGAGGTAATTTT	1620
QY	1623	TT 1624	
Db	1621	IT 1622	

RESULT 2
AF092051
LOCUS
DEFINITION Homo sapiens beta-1,3-N-acetylglucosaminyltransferase mRNA,
complete cds.
ACCESSION AF092051
VERSION AF092051.2 GI:9755416
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1831)
AUTHORS Zhou,D., Dinter,A., Gutierrez Gallego,R., Kamerling,J.P.,
Vliegthart,J.F., Berger,E.G. and Hennet,T.
TITLE A beta-1,3-N-acetylglucosaminyltransferase with
poly-N-acetylglucosamine synthase activity is structurally related

JOURNAL	to beta-1,3-galactosyltransferases
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 96 (2), 406-411 (1999)
PUBMED	99110903
REFERENCE	2 (bases 1 to 1831)
AUTHORS	Zhou,D., Berger,E.G. and Hennet,T.
TITLE	Direct Submission
JOURNAL	Submitted (15-SEP-1998) Physiology, University of Zurich, Winterthurerstrasse 190, Zurich 8057, Switzerland
REFERENCE	3 (bases 1 to 1831)
AUTHORS	Zhou,D., Berger,E.G. and Hennet,T.
TITLE	Direct Submission
JOURNAL	Submitted (09-AUG-2000) Physiology, University of Zurich, Winterthurerstrasse 190, Zurich 8057, Switzerland
REMARK	Sequence update by submitter
COMMENT	On Aug 9, 2000 this sequence version replaced gi:4191393.
FEATURES	Location/Qualifiers
source	1..1831
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/tissue_type="brain"
	/dev_stage="fetus"
	236..1429
	/EC_number="2.4.1.149"
	/note="glycosyltransferase; b3gnt"
	/codon_start=1
	/product="beta-1,3-N-acetylglucosaminyltransferase"
	/protein_id="AAD09764.2"
	/db_xref="GI:9755417"
	/translation="MSVGRRRIRKLLIMANVFIYFIMEVSKSSSOEKNKGVEIIP KEKFWKISTPPEAVNROEKLNRQYNPILSMLTNGEAGRLNISHLAVCEPDLRV TSVVTGNNLPDRKDFELLILRCNYSLLIDQPKCAKKPELLAIAISLPHFARRA IRSWGQSNAGNQTIVRVFUGQTPPDNPDLSMLKFEKRDILMMWRDPTF NLSKEVLFRLWSTSCDPTFVFKGDDDFVFNTHILNLSLSKIRAKDLFGDVI HNAQPHRDKLKYIPEVYSGLYPGGGGGLYSGHLRLVHITDQVHLPIDDV YTGCKLQKLGVLPEKHGKFTFDEEKNKNKICSYVDLMLVHRSKPKQEMIDWSQLOS AHLKC"
BASE COUNT	497 a 423 c 443 g 468 t
ORIGIN	
Query Match	89.7%; Score 1531.4; DB 9; Length 1831;
Best Local Similarity	97.5%; Pred. No. 0;
Matches 1585;	Conservative 8; Mismatches 25; Indels 7; Gaps 4;
QY	13 ACCGCGACGCGGACGCGGCAACAAAGTCCGCGAGGCTAGCAGAGCCGAGCGAGCAGT 72
Db	
QY	3 AGCGCGACGCGGACGCGGCAACAAAGTCCGCGAGGCTAGCAGAGCCGAGCGAGCAGT 62
Db	
QY	73 CCCTGCCCGCCACACCGCGCGCGCGCGCTCCGCGGCGCGCGCGATGAGCGTGAGCTGC 132
Db	
QY	63 CCCTGCCCGCCACACCGCGCGCGCGCGCTCCGCGGCGCGCGCGATGAGCGTGAGCTGC 122
Db	
QY	133 GCGGTCCGCGGCTGAGCGCGCGGAGCGCGCGGAGCTGGATGTGCGCGCGATCTCCC 192
Db	
QY	123 GCGGTCCGCGGCTGAGCGCGCGGAGCGCGCGGAGCTGGATGTGCGCGCGATCTCCC 182
Db	
QY	193 GCCTTTCGCCCGCGCGCGCGCGCGGAGCTGGATGTGCGCGCGAGCAAGATATGAGAAATGAGTG 252
Db	
QY	193 GCCTTTCGCCCGCGCGCGCGCGCGGAGCTGGATGTGCGCGCGAGCAAGATATGAGAAATGAGTG 242
Db	
QY	253 TTGAGCGTCGAAGAAATAAAGTTGTTGGGTATCCTCGATGATGGCAAAATGCTCTTCATTTATT 312
Db	
QY	243 TTGAGCGTCGAAGAAATAAAGTTGTTGGGTATCCTCGATGATGGCAAAATGCTCTTCATTTATT 302
Db	
QY	313 TTATTATGAGAGTCTCCAAAAGCAGTACCCAGAAAAAATGGAAGGGAAGTAATAA 372
Db	
QY	303 TTATTATGAGAGTCTCCAAAAGCAGTACCCAGAAAAAATGGAAGGGAAGTAATAA 362
Db	
QY	373 TACCAAGAGAGTCTCTGGAGATATCTACCCCTCCGAGGCATCTCTGGAACCGAGAGC 432
Db	
QY	363 TACCAAGAGAGTCTCTGGAGATATCTACCCCTCCGAGGCATCTCTGGAACCGAGAGC 422
Db	

Db 183 CCCCGCCCGCCGAGCTGCTCCGGACAGATATGAGAAATGAGTGTGTTGACG 242

Qy 260 TCGAAGAAATAAAGTTGTTGGTATCTCTGATGCGCAATGCTCTCATTTATTTAT 319

Db 243 TCGAAGAAATAAAGTTGTTGGTATCTCTGATGCGCAATGCTCTCATTTATTTAT 302

Qy 320 GGAAGTCTCCAAAGCAGTAGCCAAAGAAATAATGGAAGGGGAAGTATATACCCAA 379

Db 303 GGAAGTCTCCAAAGCAGTAGCCAAAGAAATAATGGAAGGGGAAGTATATACCCAA 362

Qy 380 AGAAGTCTCGAAGATATCTACCCCTCCCGAGGATAGTGAACCGGAGCAAGAGAA 439

Db 363 AGAAGTCTCGAAGATATCTACCCCTCCCGAGGATAGTGAACCGGAGCAAGAGAA 422

Qy 440 GCTGAACCGGAGTACAAACCCATCTGAGCATGCTGACCAACGAGCGGGAGCGGG 499

Db 423 GCTGAACCGGAGTACAAACCCATCTGAGCATGCTGACCAACGAGCGGGAGCGGG 482

Qy 500 CAGGCTCTCCAATATAAGCCATCTGAACCTACTCGGAACCTGACCTGAGGTCACGTCGGT 559

Db 483 CAGGCTCTCCAATATAAGCCATCTGAACCTACTCGGAACCTGACCTGAGGTCACGTCGGT 542

Qy 560 GGTACCGGTTTAAACAACTTGGCGACAGATTAAAGACTTTCTGCTGTTATGAGATG 619

Db 543 GGTACCGGTTTAAACAACTTGGCGGACAGATTAAAGACTTTCTGCTGTTATGAGATG 602

Qy 620 CCGCAATTATTACCTGCTTATAGATCAGCCGGATAGTGTGCAAGAAACCTTTCTTGT 679

Db 603 CCGCAATTATTACCTGCTTATAGATCAGCCGGATAGTGTGCAAGAAACCTTTCTTGT 662

Qy 680 GCTGGGATTAAGTCCCTCACTCCACATTTTGGCAGAAAGCAAGCAATCGGGAATCTG 739

Db 663 GCTGGGATTAAGTCCCTCACTCCACATTTTGGCAGAAAGCAAGCAATCGGGAATCTG 722

Qy 740 GGGCCAGAAAGCAACCGAGGAAACCAAGCGTGGTCCGAGTCTTCTGCTGGCCGAC 799

Db 723 GGGCCAGAAAGCAACCGAGGAAACCAAGCGTGGTCCGAGTCTTCTGCTGGCCGAC 782

Qy 800 ACCCCAGAGACAAACCCAGCCCTTTCAGATATCTGAAATTTGAGAGTGAAGCA 859

Db 783 ACCCCAGAGACAAACCCAGCCCTTTCAGATATCTGAAATTTGAGAGTGAAGCA 842

Qy 860 CCAAGACATTTATGTSAACTACAGAGACATTTCTTCAACTTGTCTCTGAAGGAAT 919

Db 843 CCAAGACATTTATGTSAACTACAGAGACATTTCTTCAACTTGTCTCTGAAGGAAT 902

Qy 920 GCTGTTCTCAGTGGGTAACTTCTCCGACACACTGAGTTGTTTTCAGGCGCA 979

Db 903 GCTGTTCTCAGTGGGTAACTTCTCCGACACACTGAGTTGTTTTCAGGCGCA 962

Qy 980 TGAGCATGTTTTGTGAACCCATCACATCCTCTGAATTTACTTGAATAGTTATCCCAAGC 1039

Db 963 TGAGCATGTTTTGTGAACCCATCACATCCTCTGAATTTACTTGAATAGTTATCCCAAGC 1022

Qy 1040 CAAAGCCAAAGATCTTTTCATAGTGTATGATCCCAATGCTGGACCTCATCGGGATA 1099

Db 1023 CAAAGCCAAAGATCTTTTCATAGTGTATGATCCCAATGCTGGACCTCATCGGGATA 1082

Qy 1100 GAAGCTGAAGTACTACATCCAGAGTTGTTTACTCTGGCTCTACCCACCTCATCGAG 1159

Db 1083 GAAGCTGAAGTACTACATCCAGAGTTGTTTACTCTGGCTCTACCCACCTCATCGAG 1142

Qy 1160 GGGAGGGGGTCTCTACTCCGCCCACCTGGCCCTGAGGCTGTACCATATCACTGACCA 1219

Db 1143 GGGAGGGGGTCTCTACTCCGCCCACCTGGCCCTGAGGCTGTACCATATCACTGACCA 1202

Qy 1220 GGTCCATCTCTACCCCATATGATGAGTTTATCTGGAATGTGCTTCAGAACTTCGGCCT 1279

Db 1203 GGTCCATCTCTACCCCATATGATGAGTTTATCTGGAATGTGCTTCAGAACTTCGGCCT 1262

Qy 1280 CGTTCAGAGAAACAAAGGCTTCAGAGACATTTGATATCGAGAGAGAAACAAATAA 1339

Db 1263 CGTTCAGAGAAACAAAGGCTTCAGAGACATTTGATATCGAGAGAGAAACAAATAA 1322

Qy 1340 CATCTGCTCTATGTAGATCTGATGTTAGTACATAGTAGAAAACCTCAAGAGATGATTGA 1399

Db 1323 CATCTGCTCTATGTAGATCTGATGTTAGTACATAGTAGAAAACCTCAAGAGATGATTGA 1382

Qy 1400 TATTTGGTCTCAGTTGCGAGAGTGCATTTAAATGCTAAATAGATACAAACTCAATTT 1459

Db 1383 TATTTGGTCTCAGTTGCGAGAGTGCATTTAAATGCTAAATAGATACAAACTCAATTT 1442

Qy 1460 KGSATWGAAGGGTWTTTTGTATWGGYCCCATGTTGGGGTCTCACATTAGAGTAATPTC 1519

Db 1443 TGCATAGAAGGTGATTTTGAATAGTTCCCATGTTGTGTTCTCACATTAGAGTAATPTC 1502

Qy 1520 TATTTNAANC--ATCAAAATTTGCTTTTATGAGTGATACCATTT-ANGGCTCTAANCCTT 1576

Db 1503 TGTATTAACCATGAAATTTGCTTTTATGAGTGATACCATTTTGGGGCTCTTAACCTT 1562

Qy 1577 ---CATTTGNACTCAGTGAAGAGGAAAGCGGAGAGGTAAPTTTTATGATGTA 1631

Db 1563 TCAATTGTTGTTACTCAGTGAAG-AGGGAAGCGGAAGATGGTAAATTTTTTTTACGGA 1619

RESULT 4

BD093191

LOCUS

DEFINITION Useful polypeptide.

ACCESSION BD093191

VERSION BD093191.1 GI:22638779

KEYWORDS WC 0100848-A/1.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1912)

AUTHORS Sasaki,K., Shiraishi,N., Natsume,A., Yamada,Y., Nakagawa,S. and Sekine,S.

TITLE Useful polypeptide

JOURNAL Patent: WO 0100848-A 1 04-JAN-2001;

COMMENT KYONA HAKKO KOGYO CO LTD,KATSUTOSHI SASAKI,NORIHIKO SHIRAISHI, AYUMI NATSUME,YOJI YAMADA,SATOSHI NAKAGAWA,SUSUMU SEKINE

OS Homo sapiens (human)

PN WO 0100848-A/1

PD 04-JAN-2001

PF 29-JUN-2000 WO 2000JP004304

PR 29-JUN-1999 JP 99P 183437,16-MAR-2000 JP 00P 074757 P1

KATSUTOSHI SASAKI,NORIHIKO SHIRAISHI,AYUMI

NATSUME,YOJI YAMADA,

P1 SATOSHI NAKAGAWA,SUSUMU SEKINE

PC C12N15/54,C12N9/10,C12N5/10,C12N1/21,A61K48/00,A61K45/00,PC A61K39/395.

PC A61K35/00,A61K31/711,A61P35/00,A61P29/00,A01K67/027,A01H5/00,

PC G01N33/53,

PC G01N33/15,G01N33/05

CC

FH Key Location/Qualifiers.

source

1..1912

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

BASE COUNT 518 a 430 c 457 g 507 t

ORIGIN

Query Match 88.1%; Score 1504.2; DB 6; Length 1912;

Best Local Similarity 97.2%; Pred.No. 0;

Matches 1582; Conservative 8; Mismatches 27; Indels 10; Gaps 6;

Qy 11 GCAGCGGAGCGGACGAGCGGCAACAGTCCGGAGGCTAGCAGAGCCGAGCA 70

Db 2 GCAGCGGAGCGGACGAGCGGCAACAGTCCGGAGGCTAGCAGAGCCGAGCA 59

Qy 71 GTCCCTGCCGCGACACCGCGGCGCCGCTCCGCGGCGCCGCGCATGGAGCGTGAGCT 130

Db 60 GTCCCTGCCGACACCGCCGCGCGCTCCGGGCGCGCGCATGGAGCGTGAGC 119
Qy 131 GCGGGGTCGCGGGGCTGAGCCCGCGGAGCGCGGCGGACGTGGATGTGGCGGATCTC 190
Db 120 GCGGGGTCGCGGGGCTGAGCCCGCGGAGC-GCCGGGACGTGGATGTGGCGGATCTC 178
Qy 191 CGGCCCTTGCCCGCCCGCGGAGCTGGAGTGTCTCCGACACAGATATGAGAATGAG 250
Db 179 CGGCCCTTGCCCGCCCGCGGAGCTGGAGTGTCTCCGACACAGATATGAGAA:GAG 238
Qy 251 TGTGGACGTCGAAGAATTAAGTGTGTGGGTATCTGTGATGCGAATGTCTTCAITTA 310
Db 239 IGTGGACGTCGAAGAATTAAGTGTGTGGGTATCTGTGATGCGAATGTCTTCAITTA 298
Qy 311 TTTTATATGGAAGTCTCCAAAGCAGTAGCCCAAGAAAAAATGAAAAAGGGAAGTAAT 370
Db 299 TTTTATATGGAAGTCTCCAAAGCAGTAGCCCAAGAAAAAATGAAAAAGGGAAGTAAT 356
Qy 371 AATACCAAGAGAGTCTCGAAGATATCTACCCCTCCGAGGCATCTGGAACCGAGA 430
Db 359 AATACCAAGAGAGTCTCGAAGATATCTACCCCTCCGAGGCATCTGGAACCGAGA 418
Qy 431 GCAAGAGAGTGAACCGGCGAGTAGAACCCCATCTGAGCATGTGACCAACGACGCG 490
Db 419 GCAAGAGAGTGAACCGGCGAGTAGAACCCCATCTGAGCATGTGACCAACGACGCG 478
Qy 491 GGAGGGGCGAGCTCTCCAATATAGCCATCTGAACCTCTGAGCATCTGAGCGGT 550
Db 479 GGAGGGGCGAGCTCTCCAATATAGCCATCTGAACCTCTGAGCATCTGAGCGGT 538
Qy 551 CACGTCGGTGTGACGGGTTTAAACACTTGGCGACAGATTTAAAGACTTTCTCTGTA 610
Db 539 CACGTCGGTGTGACGGGTTTAAACACTTGGCGACAGATTTAAAGACTTTCTCTGTA 598
Qy 611 TTGAGATCCGCATTTATCTACTGCTTATAGATCAGCGGGATAGTGTGCAAGAAC 670
Db 599 TTGAGATCCGCATTTATCTACTGCTTATAGATCAGCGGGATAGTGTGCAAGAAC 658
Qy 671 TTTCTTGTGCTGGCGATTAACTGCTCACTCCACATTTTGCAGAGCAAGCAATCG 730
Db 659 TTTCTTGTGCTGGCGATTAACTGCTCACTCCACATTTTGCAGAGCAAGCAATCG 718
Qy 731 GGATCCTGGGCGCAAGAACGACCGAGGACCAACAGGTGGTGGAGTCTTCTCT 790
Db 719 GGAATCCTGGGCGCAAGAACGACCGAGGACCAACAGGTGGTGGAGTCTTCTCT 778
Qy 791 GGCCAGACACCCAGAGACACACCCGACCTTTCAGATATCTGTAATTTGAGAG 850
Db 779 GGCCAGACACCCAGAGACACACCCGACCTTTCAGATATCTGTAATTTGAGAG 838
Qy 851 TGAGAGCACCAAGACATTTTATGTTGAACACTACAGACACATTTTCAACTTGTCT 910
Db 839 TGAGAGCACCAAGACATTTTATGTTGAACACTACAGACACATTTTCAACTTGTCT 898
Qy 911 GAGGAGTGTCTTTCTCAGGTGGTAACTTCTCTCCGACACACTGAGTTTGT 970
Db 899 GAGGAGTGTCTTTCTCAGGTGGTAACTTCTCTCCGACACACTGAGTTTGT 958
Qy 971 CAAGGGCGATGACGATGTTTGTGAACACCCATCATCTGTAATTTGTAATGATT 1030
Db 959 CAAGGGCGATGACGATGTTTGTGAACACCCATCATCTGTAATTTGTAATGATT 1018
Qy 1031 ATCAAGACCAAGACCAAGATCTTTCAATAGTGAATGATCCCAATGCTGGACCTCA 1090
Db 1019 ATCAAGACCAAGACCAAGATCTTTCAATAGTGAATGATCCCAATGCTGGACCTCA 1078
Qy 1091 TCGGGATAAGAGCTGAAGTACTACATCCGAGAGTTGTTTACTCTGGGCTCTACCCACC 1150
Db 1079 TCGGGATAAGAGCTGAAGTACTACATCCGAGAGTTGTTTACTCTGGGCTCTACCCACC 1138
Qy 1151 CTATGAGGGGAGGGGGTTCCTCTACTCCGCCACCTGGCCCTGAGGCTGTACCATAT 1210
Db 1139 CTATGAGGGGAGGGGGTTCCTCTACTCCGCCACCTGGCCCTGAGGCTGTACCATAT 1198

Qy 1211 CACTGACAGGTCCATCTCTACCCCATTTGATGACGTTTATACTGGAATGTCCTTCAGAA 1270
Db 1199 CACTGACAGAGTCCATCTCTACCCCATTTGATGACGTTTATACTGGAATGTCCTTCAGAA 1258
Qy 1271 ACTCGGCTCGTTCCAGAGAAACACAAAGGCTTCAGACATTTGATATCCAGAGAGAAA 1330
Db 1259 ACTCGGCTCGTTCCAGAGAAACACAAAGGCTTCAGACATTTGATATCCAGAGAGAAA 1318
Qy 1331 CAAAAATAACATCTGCTCTCTATGTAGATCTGTAGTGTAGTACATAGTAAAAACCTCAAGA 1390
Db 1319 CAAAAATAACATCTGCTCTCTATGTAGATCTGTAGTGTAGTACATAGTAAAAACCTCAAGA 1378
Qy 1391 GATGATTGATATTTGGTCTCAGTTCAGTTCAGTTCATTTAAATGCTAAAAATGATACAA 1450
Db 1379 GATGATTGATATTTGGTCTCAGTTCAGTTCAGTTCATTTAAATGCTAAAAATGATACAA 1438
Qy 1451 ACTCAATTTKGSATWRAAGGGTWTTTTGTATWGSYCCCATGTTGGGTCTCCACATTAG 1510
Db 1439 ACTCAATTTGCATAGAAAGGTGATTTTGAATAGTTCCTCATGTTGTCTTCACATTAG 1498
Qy 1511 AGTAATTTCTATTNAANC--ATGAAATTCCTTTATGATGATACCCATTT-ANGGCT 1567
Db 1499 AGTAATTTCTATTAAACCATGAAATTCCTTTATGATGATACCCATTTGAGGCT 1558
Qy 1568 CTAANCCCTT---CATTTGNACTCAGTGAAGAGGAAAGCGGAGAGTAAATTTTNTTT 1624
Db 1559 CTAANCCCTTCAATTTGCTACTCAGCTGAAG-AGGGAAGCGGAAGATGCTAATTTTTT 1617
Qy 1625 ATGGTGA 1631
Db 1618 TTATGGA 1624

RESULT 5
AB049584
LOCUS
DEFINITION Homo sapiens mRNA for beta-1,3-N-acetylglucosaminyltransferase
bGnt-2, complete cds.
ACCESSION AB049584
VERSION AB049584.1 GI:12619293
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
MAMMALIA; Euthera; Chordata; Craniata; Vertebrata; Euteleostomi;
Fukaryota; Metazoa; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Shiraishi,N., Natsume,A., Togayachi,A., Endo,T., Akashima,T.,
Yanada,Y., Imai,N., Nakagawa,S., Koizumi,S., Sekine,S.,
Narimatsu,H. and Sasaki,K.
TITLE Identification and characterization of three novel beta
1,3-N-acetylglucosaminyltransferases structurally related to the
beta 1,3-galactosyltransferase family
J. Biol. Chem. 276 (5), 3498-3507 (2001)
JOURNAL
MEDLINE 21264825
PMID 11042166
REFERENCE
AUTHORS Sasaki,K., Natsume,A. and Shiraishi,N.
TITLE Direct Submission
Submitted (02-OCT-2000) Katsutoshi Sasaki, The Tokyo Research
Laboratories, Kyowa Hakko Kogyo Co., Ltd., Department of Molecular
Genetics, 3-6-6 Asahi-machi, Machida-shi, Tokyo 194-8533, Japan
(E-mail:ksasaki@kyowa.co.jp, Tel:81-42-725-2555(ex.2150),
Fax:81-42-726-8330)
FEATURES
Location/Qualifiers
1. 1912
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..1912
/gene="bGnt-2"
234..1427
/gene="bGnt-2"

```
/codon_start=1
/product="beta-1,3-N-acetylglucosaminyltransferase bGnT-2"
/protein_id="BAB21530.1"
/db_xref="GI:12619294"
/translation="MSVGRRRRIKLLILMANVFIVIMEVSKSSSOENKNGKEVILP
KFKFKISTPEAYNNROEKLNRQYNPILSMLTNOTGEAGRLSLNLSHLYNCEPOLRV
TSVVTGFNLDPDFDLXLRNYSLLIDOPDKAKPFLLLAIXLSLTPHFARROA
IRESGQESNAGNOTVVRVLLGOTPPEDNHDPDLSMDLKFSEKHODILMNNYRTOF
NLSLXLYELRWVSTSCPDIEVFYKGGDDVFNTHILNLSLSKTRAKDLFIQDVI
HNAGPHRKLXIYIPEVYISGLYPYAGGGFLYSGHLALRLYHIDQVHLYPIDDV
YTGMCLOKLGVLPERKHGFRTEIDIEKNKNICSVYDEMLVHSRKPQEMIDTWSLQS
AHLKC"
BASE COUNT      518 a   430 c   457 g   507 t
ORIGIN
      1  GCGCGGCGAGCGGCGGCGGCAACAAGTGCCTGGGAGGCTAGCAGAGCGCAAGCGGAGCA 70
      2  GCGCGGCGAGCGTCACGAGCGGCAACAAGTGCCTGGGAG--TAGCAGAGCGCAAGCGGAGCA 59
      3  GTCCCTGCCCGCGACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 119
      4  GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTC 190
      5  GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTC 178
      6  CCGCCCTTGCCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 250
      7  CCGCCCTTGCCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 238
      8  TGTGACGCTGCGAAGATTAAGTTGTTGGTATCCTGATGATGGCAAAATGCTCTTCATTTA 298
      9  TTTTATATAGGAGTCTCCAAAGCAGTAGCCAAAGAAATAATGGAAGGGGGAAGTAAT 370
     10  TTTTATATAGGAGTCTCCAAAGCAGTAGCCAAAGAAATAATGGAAGGGGGAAGTAAT 358
     11  AATACCCAAAGAGAGTTCTGGAAGATATCTACCCCTCCCGAGGCGATCTGGAACCGAGA 430
     12  AATACCCAAAGAGAGTTCTGGAAGATATCTACCCCTCCCGAGGCGATCTGGAACCGAGA 418
     13  GCAAGAGAGCTGAACCGGCGAGTACAAACCCCATCTCTGAGCATGCTGACCAACCGAGCGG 490
     14  GCAAGAGAGCTGAACCGGCGAGTACAAACCCCATCTCTGAGCATGCTGACCAACCGAGCGG 478
     15  GGAGCGGCGGCGGCTCTCCAAATATAGCCATCTGAACCTAGTGGAACTGACCTGAGGCT 550
     16  GGAGCGGCGGCGGCTCTCCAAATATAGCCATCTGAACCTAGTGGAACTGACCTGAGGCT 538
     17  CAGCTCGGTGGTTACGGGTTTAAACAATCTGCGGAGCAGATTTAAAGACTTCTGCTGTA 610
     18  CAGCTCGGTGGTTACGGGTTTAAACAATCTGCGGAGCAGATTTAAAGACTTCTGCTGTA 598
     19  TTTGAGATGCGGCAATTTACTGCTTATAGATCAGCCCGATAGTGTGCAAGAAACCC 670
     20  TTTGAGATGCGGCAATTTACTGCTTATAGATCAGCCCGATAGTGTGCAAGAAACCC 658
     21  TTTCTTGTGCTGCGGATTAAGTCCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 730
     22  TTTCTTGTGCTGCGGATTAAGTCCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 718
     23  GGAATCTGCGGCGCAAGAAAGCAACGAGGAGCAACCGTGGTGGAGTCTTCTGCT 790
     24  GGAATCTGCGGCGCAAGAAAGCAACGAGGAGCAACCGTGGTGGAGTCTTCTGCT 778
     25  GGGCGAGACACCCCGGAGGACCAACCCCGGAGCTTTCAGATATGCTGAATTTGAGAG 850
```

```
779 GGGCCAGACACCCCGAGAGCAACACCCCGACCTTTTCAGATATGCTGAAATTTGAGAG 838
851 TGAGAAGCACCAAGACATTTCTATGCTGGAATACAGAGACACTTTTCTCAACTTGTCTCT 910
839 TGAGAAGCACCAAGACATTTCTATGCTGGAATACAGAGACACTTTTCTCAACTTGTCTCT 898
911 GAAGGAAGTGTCTTCTCAGTGGGTAAAGTACTTCTCCGCCAGACACTGAGTTTGTGTTT 970
899 GAAGGAAGTGTCTTCTCAGTGGGTAAAGTACTTCTCCGCCAGACACTGAGTTTGTGTTT 958
971 CAAGCGCATGACGATGTTTGTGGAACACCCATCACATCTGAAATTAATCTTGAATAGTTT 1030
959 CAAGCGCATGACGATGTTTGTGGAACACCCATCACATCTGAAATTAATCTTGAATAGTTT 1018
1031 ATCCAGACCAAGCAAGATCTCTCATAGTGTGATGATCCACATGCTGGACCTCA 1090
1019 ATCCAGACCAAGCAAGATCTCTCATAGTGTGATGATCCACATGCTGGACCTCA 1078
1091 TCGGATTAAGAGCTGAAAGTACTACATCCAGAAAGTGTGTTTACTCTGGCCTTACCCACC 1150
1079 TCGGATTAAGAGCTGAAAGTACTACATCCAGAAAGTGTGTTTACTCTGGCCTTACCCACC 1138
1151 CTATCAGGGGAGGGGGTTCCTCTACTCGGCCACCTGGCCCTGAGGCTGTACCATAT 1210
1139 CTATCAGGGGAGGGGGTTCCTCTACTCGGCCACCTGGCCCTGAGGCTGTACCATAT 1198
1211 CACTCAGAGGTCATCTTACCCATTTGATGACCTTTATCTGGAATGTGCTTCAGAA 1270
1199 CACTCAGAGGTCATCTTACCCATTTGATGACCTTTATCTGGAATGTGCTTCAGAA 1258
1271 ACTCGGCTCGTTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATCGAGGAGAAA 1330
1259 ACTCGGCTCGTTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATCGAGGAGAAA 1318
1331 CAAAATTAACATCTCTCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1390
1319 CAAAATTAACATCTCTCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1378
1391 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1450
1379 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1438
1451 ACTCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1510
1439 ACTCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1498
1511 AGTAATTTCTATTNANC--ATGAAATGCTTATGATGATGATGATGATGATGATGATGATGAT 1567
1499 AGTAATTTCTATTAAACCATGAAATTTGCTTTATGATGATGATGATGATGATGATGATGAT 1558
1568 CTAACCTTT--CATTTGNACTCAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1624
1559 CTAACCTTTCAATTTGGTACTCAGTGAAG--AGGAAAGCGGAGGAGGAGGAGGAGGAGGAG 1617
1625 ATGGTGA 1631
1618 TTATGGA 1624
```

RESULT 6
AC093401
LOCUS
DEFINITION
AC093401
VERSION
AC093401.4
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 99089)
AUTHORS
Sulston,J.E. and Waterston,R.
TITLE
Toward a complete human genome sequence

JOURNAL

99063792
MEDLINE
PUBMED

REFERENCE

2 (bases 1 to 99088)
Holmes, A. and Cotton, M.

AUTHORS

The sequence of Homo sapiens BAC clone RP11-93M19
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

JOURNAL

3 (bases 1 to 99088)
Waterston, R. H.

AUTHORS

Direct Submission

TITLE

Submitted (22-AUG-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE

4 (bases 1 to 99088)
Waterston, R. H.

AUTHORS

Direct Submission

TITLE

Submitted (20-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE

5 (bases 1 to 99088)
Waterston, R. H.

AUTHORS

Direct Submission

TITLE

Submitted (29-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE

6 (bases 1 to 99088)
Waterston, R.

AUTHORS

Direct Submission

TITLE

Submitted (16-APR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 29, 2002 this sequence version replaced gi:19424712.

COMMENT

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_NH0093M19

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPC1-11 human BAC library was made from the blood of one male
donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frengen, E.,
Tateno, M., Catanese, J. J. and de Jong, P. J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong
and coworkers at <http://www.chori.org>
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-270B14, 2000 bp overlap;
the clone sequenced to the right is RP11-642B6, 2000 bp overlap.
Actual end of this clone is at base position 38298 of RP11-642B6.

Sequence derived from one M13 subclone, base position 30875 to
30886.

Polymorphisms have been identified between AC018462, AC093159 and
AC093401.

FEATURES

Location/Qualifiers

SOURCE

1..99088
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"

/clone="RP11-93M19"

/clone_lib="RPC1-11"

repeat_region

218..337
/rpt_family="MIR"

repeat_region

906..1160
/rpt_family="Alu"

repeat_region

1379..1925
/rpt_family="ERV1"

repeat_region

2261..2297
/rpt_family="MIR"

repeat_region

2571..2939
/rpt_family="L1"

repeat_region

2974..3081
/rpt_family="MIR"

repeat_region

3123..3225
/rpt_family="Alu"

repeat_region

3486..3708
/rpt_family="MIR"

misc_feature

3789..4320
/note="match to EST BG721682 (NID:g14000869)"

misc_feature

3829..4087
/note="match to EST AA377082 (NID:g2029630)"

misc_feature

3869..4691
/note="match to EST AU137528 (NID:g10998067)"

misc_feature

4100..4563
/note="match to EST H78875 (NID:g1056964) yu09g06.rl"

misc_feature

4141..4441
/note="match to EST AA263148 (NID:g1898954)"

misc_feature

4142..4793
/note="match to EST BG741357 (NID:g14052010)"

misc_feature

4180..4678
/note="similar to Mus musculus EST AA204363 (NID:g1801007)
mu30ell.rl"

misc_feature

4264..4724
/note="match to EST AW958548 (NID:g8148232)"

misc_feature

4426..5325
/note="similar to EST BM557603 (NID:g18799721)"

misc_feature

4468..4938
/note="similar to Mus musculus EST BE632125 (NID:g9914895)
uu57d09.yl"

misc_feature

4498..4879
/note="match to EST H47990 (NID:g924042) yp75dll.rl"

misc_feature

4624..5231
/note="match to EST BF001654 (NID:g10701929)"

misc_feature

4666..5089
/note="match to EST AV682596 (NID:g10284459)"

misc_feature

4680..5391
/note="match to EST BE622082 (NID:g9893022)"

misc_feature

4915..5257
/note="match to EST R31722 (NID:g787565) yhe3d09.rl"

misc_feature

4945..5390
/note="match to EST H13125 (NID:g877945) yj06e09.rl"

misc_feature

4948..5597
/note="match to EST BF131058 (NID:g10970087)"

misc_feature

5026
/note="match to EST R74552 (NID:g848922) yi58a02.rl"

misc_feature

5031..5682
/note="match to EST BE564288 (NID:g9808008)"

misc_feature

5061..5553
/note="match to EST N58174 (NID:g1202064) yv6a07.s1"

LOCUS BC030579 2454 bp mRNA linear PRI 21-MAY-2002
DEFINITION Homo sapiens, UDP-GlcNAc:betaGal
beta-1,3-N-acetylglucosaminyltransferase 1, clone MGC:26071
IMAGE:4828158, mRNA, complete cds.
ACCESSION BC030579
VERSION 1 GI:21040508
SOURCE MGC.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2454)
Direct Submission
Submitted (13-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgaps-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 34 Row: d Column: 20
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 15451893.
Location/Qualifiers
1..2454
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="LocusID:10678"
/db_xref="taxon:9606"
/clone="MGC:26071 IMAGE:4828158"
/tissue_type="Testis"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"
113..1306
/codon_start=1
/product="UDP-GlcNAc:betaGal
beta-1,3-N-acetylglucosaminyltransferase 1"
/protein_id="AAH30579.1"
/db_xref="GI:21040509"
/translation="MSVGRRIKLLGLMANVFIYFIMEVSKSSQKNGKEVILP
KEKFWKISPPPEYHNREGEKLNQNFILSMILNOTGEARLSNLSHLYNCEPDLRV
TSVTFNPNLRFKDFLLYLRCRNSYLLIDQDKAKPFLILAKLSLIPFARQA
IRESQESNAGNOTVYRVFLIGQTPEDNHPDLSDMLKPFSEKHQDILMMYRDTFF
NLSKVELFLRWVTSCTPTEFVFKGDDVFVNTHLNLYNSLSTKAKDLFIQGV
HNAGPHRDKLKYIPEVYVSGLYPPYAGGGFLYSGHLALRLYHITDOVHLYPIDDV
YTMQCLQKLGVLPEKHKGFRTDIEKNKNKNCISYVDLMVLSRKPQEMID:WSLQLS
AHLKC"
BASE COUNT 725 a 444 c 524 g 761 t
ORIGIN
Query Match 76.7%; Score 1309.4; DB 9; Length 2454;
Best Local Similarity 97.1%; Pred. No. 1.2e-293;
Matches 1363; Conservative 8; Mismatches 25; Indels 7; Gaps 4;
Qy 235 AGATATGAGAATGAGTGTGGACGTCGAAGATAAAGTTGTGGGTATCTTCATGATGCG 294
|||||
Db 102 AGATATGAGAATGAGTGTGGACGTCGAAGATAAAGTTGTGGGTATCTTCATGATGCG 161
Qy 295 CAAATGCTTCATTTATTTATGGAAGTCTCCAAAAGCAGTAGCCCAAGAAAAAATG 354
|||||

|||||
162 CAAATGCTTCATTTATTTATGGAAGTCTCCAAAAGCAGTAGCCCAAGAAAAAATG 221
Qy 355 GAAAGGGGAGTAAATAATACCCAAAGAGAGTTCIGGAAGATATCTACCCCTCCCGAGG 414
|||||
222 GAAAGGGGAGTAAATAATACCCAAAGAGAGTTCIGGAAGATATCTACCCCTCCCGAGG 281
Qy 415 CATACTGGAACCGAGAGCAAGAGAGCTGAACCGGCAGTCAACCCCATCTCIGAGCATGC 474
|||||
282 CATACTGGAACCGAGAGCAAGAGAGCTGAACCGGCAGTCAACCCCATCTCIGAGCATGC 341
Qy 475 TGACCAACAGACGCGGGGAGGCGGCGCTCTCCAATATAAGGCATCTGAACTACTGCG 534
|||||
342 TGACCAACAGACGCGGGGAGGCGGCGCTCTCCAATATAAGGCATCTGAACTACTGCG 402
Qy 535 AACCTGACCTCAGGCTCAGCTCGGTGGTTCAGGGTTTAAACAACCTGCCGACAGATTTA 594
|||||
402 AACCTGACCTCAGGCTCAGCTCGGTGGTTCAGGGTTTAAACAACCTGCCGACAGATTTA 461
Qy 595 AAGACTTTCGTCTGTTTGGAGTGCCTGCAATATTATTCACCTGCTTATAGATCAGCGGATA 654
|||||
462 AAGACTTTCGTCTGTTTGGAGTGCCTGCAATATTATTCACCTGCTTATAGATCAGCGGATA 521
Qy 655 AGTGTGCAAGAAACCTTCTTGTGTGGCGATTAGTCCCTCACTCCACATTTTGGCA 714
|||||
522 AGTGTGCAAGAAACCTTCTTGTGTGGCGATTAGTCCCTCACTCCACATTTTGGCA 581
Qy 715 GAAGCAAGCAATCGGGAATCTTGGGGCAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAG 774
|||||
582 GAAGCAAGCAATCGGGAATCTTGGGGCAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAG 641
Qy 775 TCGAGTCTTCTGCTGGCGCAGACACCCAGAGGACCAACCCAGCCACCTTTCAGATA 834
|||||
642 TCGAGTCTTCTGCTGGCGCAGACACCCAGAGGACCAACCCAGCCACCTTTCAGATA 701
Qy 835 TGCTGAATTTGAGAGTGAAGAACCAAGACATTTCTTATGTGGAACCTACAGACACTT 894
|||||
702 TGCTGAATTTGAGAGTGAAGAACCAAGACATTTCTTATGTGGAACCTACAGACACTT 761
Qy 895 TCTTCAACTTCTCTGCTGAAGGAAGTCTTCTCAGGTGGTGAAGTACTTCTCGCCGAG 954
|||||
762 TCTTCAACTTCTCTGCTGAAGGAAGTCTTCTCAGGTGGTGAAGTACTTCTCGCCGAG 821
Qy 955 ACACAGTGTGTTGTTTCAAGGGCGATGACGATGTTTGTGAACACCCATCACATCTCTGA 1014
|||||
822 ACACAGTGTGTTGTTTCAAGGGCGATGACGATGTTTGTGAACACCCATCACATCTCTGA 881
Qy 1015 ATTACTTGAATAGTTTATCCAGAGCAAGCCAAAGATCTCTTATAGGTGATGTATCC 1074
|||||
882 ATTACTTGAATAGTTTATCCAGAGCAAGCCAAAGATCTCTTATAGGTGATGTATCC 941
Qy 1075 ACAATGCTGGACCTCATCGGGATAAGAGCTCAAGTACTACATCCCAAGTGTGTTTACT 1134
|||||
942 ACAATGCTGGACCTCATCGGGATAAGAGCTCAAGTACTACATCCCAAGTGTGTTTACT 1001
Qy 1135 CTGGCTCTACCCACCTTATCGAGGGGAGGGGGTCTCTTACTCCGGCCACCTGGGCC 1194
|||||
1002 CTGGCTCTACCCACCTTATCGAGGGGAGGGGGTCTCTTACTCCGGCCACCTGGGCC 1061
Qy 1195 TGAGGCTGTACCATATCACTGACAGGTCATCTCTACCCCATGTGATGACGTTTATCTG 1254
|||||
1062 TGAGGCTGTACCATATCACTGACAGGTCATCTCTACCCCATGTGATGACGTTTATCTG 1121
Qy 1255 GAATGCTGCTTCAGAAACTCGGCTCGTTCAGAGAGAACCAAGGCTTCAGGACATTTG 1314
|||||
1122 GAATGCTGCTTCAGAAACTCGGCTCGTTCAGAGAGAACCAAGGCTTCAGGACATTTG 1181
Qy 1315 ATATCGAGGAGAAAAACAAAATAACATCTGCTCTCTATGTAGATCTGATGTAGTACATA 1374
|||||
1182 ATATCGAGGAGAAAAACAAAATAACATCTGCTCTCTATGTAGATCTGATGTAGTACATA 1241
Qy 1375 GTAGAAAAACCTCAAGAGATGATTGATTTGGTCTCAGTTCAGAGTGCTCATTTAAAT 1434
|||||

Db 1242 GTAGAAACCTCAGAGATGATTGATATTGGTCTCAGTGCAGAGTGCTCATTTAAAT 1301

Qy 1435 GCTAAATAGATACAACTCAATTTTGSATWRAAGGGGTWTTTGTATWGYCCCATGT 1494
|||||
Db 1302 GCTAAATAGATACAACTCAATTTTGCATAGAAAGGTGTATTTTGAATAGTCCCATGT 1361

Qy 1495 TGGGGTCTCATTAGAGTAATTTCTATTNANC--ATGAATTCCTTTTATGAGTAT 1552
|||
Db 1362 TGTGTTCTCATTAGAGTAATTTCTATTNANC--ATGAATTCCTTTTATGAGTAT 1421

Qy 1553 ACCCATTT-ANGGCTCTAANCCTT---CATTTGNACTCAGCTGAAGAAGGGAACGGG 1608
|||||
Db 1422 ACCCATTTGAGGGCTCTAACCCTTCAATTTGGTACTCAGCTGAAG-AGGGAAGCGGA 1480

Qy 1609 AGAAGGTAAATTTTATGATGTA 1631
|||
Db 1481 AGATGGTAATTTTTTTTACGGA 1503

RESULT 8
AF288209 2516 bp mRNA linear PRI 03-AUG-2000
LOCUS Homo sapiens beta galactosyltransferase bGalt7 mRNA, complete cds.
DEFINITION AF288209
ACCESSION AF288209
VERSION AF288209.1 GI:9664888
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2516)
AUTHORS Gromova,I., Gromov,P. and Celis,J.
TITLE A novel member of beta-1,3-galactosyltransferase family is down
regulated during bladder TCC progression
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2516)
AUTHORS Gromova,I., Gromov,P. and Celis,J.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-2000) Medical Biochemistry, University of Aarhus,
Ole Worms Alle 170, Aarhus C-8000, Denmark
FEATURES
source Location/Qualifiers
1..2516
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="bladder cancer"
43..1224
/product="beta galactosyltransferase bGalt7"
/protein_id="AAF97254.1"
/db_xref="GI:9664889"
/translation="MVSRSLVGLIMANVFIFIMEVSKSSSQEKNKGVEIIPKEKF
WKISTPEAYNPEQEKLNROYNPILSMLTNQTGEAGRLSNISHLNCEPDLRV-SVY
TGFNLPDRFKDFLLYLRGNYSLLIDQPKCAKPFLLAIKSLPHFARQAIRES
WGSEHAGNQTVPVRFELGTPPDHNPDLSDMLKFESEKHODILMWNTRDFNLSL
KEYFLRWSTYDFVFKGDDVFNTHHILNLSKTKAKDLFIDGVHINAG
PHRDKLYIPEVYSGLYPYAGGGGLYSGLHALRLYHITDQVHLYPDIDVYTG
CLQKGLVPEKHKGRTFDIEKNKNICSYVDLMLVSHKPKQEMIDFISQLQSAHLK
C*"

BASE COUNT 745 a 440 c 522 g 809 t

ORIGIN

Query Match 74.4% Score 1269.8; DB 9; Length 2516;
Best Local Similarity 97.0%; Pred. No. 2e-284;
Matches 1324; Conservative 8; Mismatches 26; Indels 7; Gaps 4;

Qy 273 TTGTTGGGTATCTGATGATGGCAATGTCTTCATTATTTTATGGAAGTCTCCAA 332
|||
Db 58 TTGTTGGGTATCTGATGATGGCAATGTCTTCATTATTTTATGGAAGTCTCCAA 117

Qy 333 AGCAGTAGCCCAAGAAAAAATGGAAGGGGAAGTAATAATACCAAGAGAAGTCTCG 392
|||||
Db 118 AGCAGTAGCCCAAGAAAAAATGGAAGGGGAAGTAATAATACCAAGAGAAGTCTCG 177

Qy 393 AAGATATCTACCCCTCCGAGGCGATACTGGAACCGAGAGCAAGAGCTGAACCGCGAG 452
|||||
Db 178 AAGATATCTACCCCTCCGAGGCGATACTGGAACCGAGAGCAAGAGCTGAACCGCGAG 237

Qy 453 TACAAACCCATCTCTGAGCATGCTGACCAACAGAGCGGGAGGGCGGCTCTCCAAT 512
|||||
Db 238 TACAAACCCATCTCTGAGCATGCTGACCAACAGAGCGGGAGGGCGGCTCTCCAAT 297

Qy 513 ATAAGCCATCTGAATCTGGAACCTGACCTGAGGGTTCAGCTGGTGGTTTACGGGTTT 572
|||||
Db 298 ATAAGCCATCTGAATCTGGAACCTGACCTGAGGGTTCAGCTGGTGGTTTACGGGTTT 357

Qy 573 AACAACTTGGCGGACAGATTTAAAGACTTCTGCTGTATTGAGATGCCGCAATATTC 632
|||||
Db 358 AACAACTTGGCGGACAGATTTAAAGACTTCTGCTGTATTGAGATGCCGCAATATTC 417

Qy 633 CTGCTTATAGATCAGCGGATTAAGTGTCAAAGAAACCTTTCTTGTGCTGGCATTAA 692
|||||
Db 418 CTGCTTATAGATCAGCGGATTAAGTGTCAAAGAAACCTTTCTTGTGCTGGCATTAA 477

Qy 693 TCCCTACTCCACATTTTGGCAGAGGAAGCAATCCGGGAATCCTGGGGCCAAAGAAC 752
|||||
Db 478 TCCCTACTCCACATTTTGGCAGAGGAAGCAATCCGGGAATCCTGGGGCCAAAGAAC 537

Qy 753 AACCGAGGAACCAACCGTGTGGAGTCTTCTGCTGGGCCAGACACCCCGAGAGAC 812
|||||
Db 538 AACCGAGGAACCAACCGTGTGGAGTCTTCTGCTGGGCCAGACACCCCGAGAGAC 597

Qy 813 AACCAACCCGACCTTTCAGATATGCTGAAATTTGAGAGTGAGAAGCACCAAGACATCT 872
|||||
Db 598 AACCAACCCGACCTTTCAGATATGCTGAAATTTGAGAGTGAGAAGCACCAAGACATCT 657

Qy 873 ATGTGGAACTACAGAGACACTTTCTTCAACTGTCTCTGAAGGAAGTCTGTTTCTCAG 932
|||||
Db 658 ATGTGGAACTACAGAGACACTTTCTTCAACTGTCTCTGAAGGAAGTCTGTTTCTCAG 717

Qy 933 TGGGTAAGTACTTCTGCCAGACACTGAGTTGTTTTCAGAGGGCGATGACGATGTTT 992
|||||
Db 718 TGGGTAAGTACTTCTGCCAGACACTGAGTTGTTTTCAGAGGGCGATGACGATGTTT 777

Qy 993 GTGAACACCCATCAGATCCTGAAATTTAGTATTTATCCAGACCAAGACCAAGAT 1052
|||||
Db 778 GTGAACACCCATCAGATCCTGAAATTTAGTATTTATCCAGACCAAGACCAAGAT 837

Qy 1053 CTCCTCATAGGTGATGATCCACAATGCTGACCTCATCGGATTAAGAGCTGAAGTAC 1112
|||||
Db 838 CTCCTCATAGGTGATGATCCACAATGCTGACCTCATCGGATTAAGAGCTGAAGTAC 897

Qy 1113 TACATCCCAAGAGTTGTTTACTCTGGCCTCTACCCACCTATGACAGGGGAGGGGGTTC 1172
|||||
Db 898 TACATCCCAAGAGTTGTTTACTCTGGCCTCTACCCACCTATGACAGGGGAGGGGGTTC 957

Qy 1173 CTCCTACTCGGCCACTCGCCCTGAGGCTGTACCATATCTACTGACAGGTCTCATCTAC 1232
|||||
Db 958 CTCCTACTCGGCCACTCGCCCTGAGGCTGTACCATATCTACTGACAGGTCTCATCTAC 1017

Qy 1233 CCCATTGATGAGCTTTATATGGAATGTGCCTTCCAGAACTCGGCCTGTTCCAGAGAAA 1292
|||||
Db 1018 CCCATTGATGAGCTTTATATGGAATGTGCCTTCCAGAACTCGGCCTGTTCCAGAGAAA 1077

Qy 1293 CACAAAGGCTTCAGGACATTTGATATCGAGGAGAAAAACAAAAATACATCTGCTCCTAT 1352
|||||
Db 1078 CACAAAGGCTTCAGGACATTTGATATCGAGGAGAAAAACAAAAATACATCTGCTCCTAT 1137

Qy 1353 GTAGATCTGATGTTAGTACATAGTAGAAACCTCAAGAGATGATGATATTGGTCTCAG 1412
|||||
Db 1138 GTAGATCTGATGTTAGTACATAGTAGAAACCTCAAGAGATGATGATATTGGTCTCAG 1197

Qy 1413 TTGCAGAGTGCTCATTTTAAATAGTAAATAGATACAAACTCAATTTTGGTATWGAAGG 1472
|||||
Db 1198 TTGCAGAGTGCTCATTTTAAATAGTAAATAGATACAAACTCAATTTTGGTATWGAAGG 1257

Qy	1473	GTWTTTGRATWGYCCCATGTTGGGGTCTCACAAITAGAGTAATTTCTATTINAAGC--A	1536
Db	1258	GTATTTTGAATAGTTCOCATGTTGTGTTCTCACATTAGAGTAATTTCTGATTAAACCAT	1317
Qy	1531	TGAAATTCGCTTTATGAGTCATACCCATTTC--ANGGCCICTAANCCCTT--CATTTGNACT	1586
Db	1318	GAAATTCGCTTTATGAGTGATACCCANTTTGAGGGCCTCTAAACCTTCCTCAATTGGTACT	1377
Qy	1587	CACGTGAAGGAGGAAACCGGAGGTAATTTTATGGTGA	1631
Db	1378	CACGTGAAG-AGGCAACCGAAGATGCTAATTTT--TTTACGGA	1421
<p>RESULT 9</p> <p>BD156964</p> <p>LOCUS</p> <p>DEFINITION</p> <p>ACCESSION</p> <p>VERSION</p> <p>KEYWORDS</p> <p>SOURCE</p> <p>ORGANISM</p>			
<p>2442 bp DNA linear PAT 17-JAN-2003</p> <p>Primer for synthesizing full-length cDNA and use thereof.</p>			
<p>BD156964</p> <p>BD156964.1 GI:27862722</p> <p>JP 2002191363-A/11807.</p> <p>Homo sapiens (human)</p> <p>Homo sapiens</p>			
<p>REFERENCE</p> <p>1 (bases 1 to 2442)</p> <p>Authors</p> <p>Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Ishii,S., Sugiyama,I., Wakamatsu,A., Nagai,K. and Otsuki,T.</p> <p>Title</p> <p>Primer for synthesizing full-length cDNA and use thereof</p> <p>Journal</p> <p>Patent: JP 2002191363-A 11807 09-JUL-2002;</p> <p>COMMENT</p> <p>OS Homo sapiens (human)</p> <p>PN JP 2002191363-A/11807</p> <p>PD 09-JUL-2002</p> <p>PF 28-JUL-2000 JP 2000280990</p> <p>PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU SAITO,</p> <p>PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU, PI KEIICHI NAGAI, TETSUJI OTSUKI</p> <p>PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC 10,</p> <p>PC C12P21/02,C1201/58/C12P21/08,C06F17/30,C12N15/00,C12N5/00 CC</p> <p>Primer for synthesizing full-length cDNA and use thereof FH Key</p>			
<p>FT CDS</p> <p>Location/Qualifiers</p> <p>1..2442</p> <p>/organism="Homo sapiens"</p> <p>/mol_type="genomic DNA"</p> <p>/db_xref="taxon:9606"</p>			
<p>BASE COUNT</p> <p>717 a 433 c 503 g 789 t</p> <p>ORIGIN</p>			
<p>Query Match 72.1%; Score 1231.4; DB 6; Length 2442;</p> <p>Best Local Similarity 97.0%; Pred. No. 1.7e-275;</p> <p>Matches 1285; Conservative 8; Mismatches 25; Indels 7; Gaps 4;</p>			
Qy	313	TTATTATGGAAGTCTCCAAAGCAGTAGTACCAAGAAAAAATGGAAGGGGGAAGTATAAA	372
Db	1	TTATTATGGAAGTCTCCAAAGCAGTAGTACCAAGAAAAAATGGAAGGGGGAAGTATAAA	60
Qy	373	TACCAAGAGAAAGTCTCGAAGATATCTACCCCTCCCGAGGCATCTCGAACCCGAGAGC	432
Db	61	TACCAAGAGAAAGTCTCGAAGATATCTACCCCTCCCGAGGCATCTCGAACCCGAGAGC	120
Qy	433	AAGAGAAGCTGAACCGGCGAGTACAAACCCCATCTCGAGCATGCTGACCAACGACGCGGGG	492
Db	121	AAGAGAAGCTGAACCGGCGAGTACAAACCCCATCTCGAGCATGCTGACCAACGACGCGGGG	180
Qy	493	AGCGGCGAGGCTCTCCAATATAGCCATCTGAAGTACTCGAACCTGACCTGAGGCTCA	552
Db	181	AGCGGCGAGGCTCTCCAATATAGCCATCTGAAGTACTCGAACCTGACCTGAGGCTCA	240

```
Db      1320  ACGA 1324
||
RESULT 10
AK002009
LOCUS   Homo sapiens cDNA FLJ11147 fis, clone PLACE1006678, weakly similar
DEFINITION
ACCESSION AK002009
VERSION   Homo sapiens mRNA for type II membrane protein, clone:HP10328.
KEYWORDS  oligo capping; fis (full insert sequence).
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens

REFERENCE
AUTHORS  Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Ishibashi,T., Fujimori,K.,
Tanai,H., Kimata,M., Watanabe,M., Hiraoka,S., Ishii,S., Kawai,Y.,
Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Negahari,K.,
Masuno,Y. and Kanehori,K.
JOURNAL  NEDO human cDNA sequencing project
REFERENCE 2 (bases 1 to 2442)
AUTHORS  Isogai,T. and Otsuki,T.
TITLE    Direct Submission
JOURNAL  Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT  NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction; 5'- & 3'-end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
source
    Location/Qualifiers
        1..2442
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="PLACE:006678"
            /tissue_type="placenta"
            /clone_lib="PLACE1"
            /note="cloning vector: pME18SFL3"
        6..1127
            /note="unnamed protein product."
            /codon_start=1
            /protein_id="BAA92031.1"
            /db_xref="GI:7023630"
            /translation="MEVSKSSSQEKNKGKEVIIPKEFKWIKISIPPEAVWNRQEKLNK
QYNPLSLMTNQTGEARLSNLSHLYCEPDLRYTSVYTGNNLPDRKDFLLYLCR
NYSLLIDPDCAKPELLLAIKSLTFHARRQAIRESWGOESNAGNQTVVRVFLGG
TPPNDHPDLSMDLKFSEKHKQDILMNRYDFTFNLSKLVFLRWSTSCPTFEVY
KGDDVFVNTTHILNLSLTKAKDLFIGDVTHNAGPHRDKKLYIIEVYVSGLY
PPYAGGGFLGSHLALRYHITDQVHLYPIDQVYTGMCLOGLVPEKHKGFRTFDI
EKNKNNTICSVVDLMVLVHSRKPQEMDIWSLQSAHLKC"
BASE COUNT 717 a 433 c 503 g 789 t
ORIGIN
Query Match 72.1%; Score 1231.4; DB 9; Length 2442;
Best Local Similarity 97.0%; Pred. No. 1.7e-275;
Matches 1285; Conservative 8; Mismatches 25; Indels 7; Gaps 4;

QY      313  TTATTATGAAGTCTCCAAAAGCAGTACGCCAAGAAAAAATGGAAAAGGGGAAGTAATAA 372
Db      1  TTATTATGAAGTCTCCAAAAGCAGTACGCCAAGAAAAAATGGAAAAGGGGAAGTAATAA 60

QY      373  TACCAAAAGAGAAGTCTTGGAGATATCTACCCCTCCCGAGGCATCTGGGAACCGAGAGC 432
Db      61  TACCAAAAGAGAAGTCTTGGAGATATCTACCCCTCCCGAGGCATCTGGGAACCGAGAGC 120

433  AACGAAAGCTGAACCGGCAGTACACCCCATCTCTGAGCATCTGTGACCAACCGACGCGGG 492
121  AAGAAAGCTGAACCGGCAGTACACCCCATCTCTGAGCATCTGTGACCAACCGACGCGGG 180
493  AGCGGGCAGGCTCTCCAATATAAGCCATCTGAACTACTGCGAACCTGACCTGAGGTCGA 552
181  AGCGGGCAGGCTCTCCAATATAAGCCATCTGAACTACTGCGAACCTGACCTGAGGTCGA 240
553  CGTCGGTGGTTAGGGGTTTAAACAACCTTCCGGACAGATTTAAAGACTTTCTGCTGTATT 612
241  CGTCGGTGGTTAGGGGTTTAAACAACCTTCCGGACAGATTTAAAGACTTTCTGCTGTATT 300
613  TGAGATGCCGCAATTATTCACCTGCTTATAGATCAGCCGATAGTGTCGCAAGAACCTT 672
301  TGAGATGCCGCAATTATTCACCTGCTTATAGATCAGCCGATAGTGTCGCAAGAACCTT 360
673  TCTTGTGTGGCGATTAAAGTCCCTCACCTCCACATTTTCCAGAGGCAAGCAATCCGGG 732
361  TCTTGTGTGGCGATTAAAGTCCCTCACCTCCACATTTTCCAGAGGCAAGCAATCCGGG 420
733  AATCCTGGGGCCAAAGAAAGCAAGCGAGGAAACCAACGGTGGTGGAGTCTTCTGCTGG 792
421  AATCCTGGGGCCAAAGAAAGCAAGCGAGGAAACCAACGGTGGTGGAGTCTTCTGCTGG 480
793  GCCAGACACCCCCAGAGGACCAACCCCGACCTTTTCAGATATGCTGAAATTTGAGAGTG 852
481  GCCAGACACCCCCAGAGGACCAACCCCGACCTTTTCAGATATGCTGAAATTTGAGAGTG 540
853  AGAAGCACCACAGACATCTTATGTGGAATACAGACACATTTCTTCAACTTGTCTCTGA 912
541  AGAAGCACCACAGACATCTTATGTGGAATACAGACACATTTCTTCAACTTGTCTCTGA 600
913  AGAAGTGTCTTCTCAGTGGTAAAGTACTTCTCGCCAGACACTGAGTTGTTTTCGA 972
601  AGAAGTGTCTTCTCAGTGGTAAAGTACTTCTCGCCAGACACTGAGTTGTTTTCGA 660
973  AGGGCGATGACGATGTTTTTGTGAACACCCATCAGATCTGAAATTTACITGAAATAGT 1032
661  AGGGCGATGACGATGTTTTTGTGAACACCCATCAGATCTGAAATTTACITGAAATAGT 720
1033  CCAAGACCAAGCCAAAGATCTTTCATAGTGAITGTATCCACAATGCTGGACCTCATC 1092
721  CCAAGACCAAGCCAAAGATCTTTCATAGTGAITGTATCCACAATGCTGGACCTCATC 780
1093  GGATAGAAGCTGAAGTACTACATCCAGAGTCTTACTCTGGCTCTACCCACCT 1152
781  GGATAGAAGCTGAAGTACTACATCCAGAGTCTTACTCTGGCTCTACCCACCT 840
1153  ATGACGGGGAGGGGGTTCCTCTACTCCGGCCACCTGGCCCTGAGGCTGTACCATATCA 1212
841  ATGACGGGGAGGGGGTTCCTCTACTCCGGCCACCTGGCCCTGAGGCTGTACCATATCA 900
1213  CTGACCAAGTCCATCTCTACCCCATGATGACGCTTTTACTGGAATGTGCCITCAGAAC 1272
901  CTGACCAAGTCCATCTCTACCCCATGATGACGCTTTTACTGGAATGTGCCITCAGAAC 960
1273  TCGGCTCTGTTCCAGAGAAACAAAGGCTTCAGGACATTTGATATCGAGGAGAAAAACA 1332
961  TCGGCTCTGTTCCAGAGAAACAAAGGCTTCAGGACATTTGATATCGAGGAGAAAAACA 1020
1333  AAAATAACATCTCTCTCTATGTAGATCTGTATGTAGTACATAGTAAACCTCAAGAGA 1392
1021  AAAATAACATCTCTCTCTATGTAGATCTGTATGTAGTACATAGTAAACCTCAAGAGA 1080
1393  TGATTGATATTGGTCTCAGTTCAGAGTCTCATTTTAAATGCTAAATAGATACAAAC 1452
1081  TGATTGATATTGGTCTCAGTTCAGAGTCTCATTTTAAATGCTAAATAGATACAAAC 1140
1453  TCATTTKGSATWGRAGGGTWTTTTGRATWGGYCCCATGTTGGGCTCTCACATTAGAG 1512
1141  TCAATTTTGCATGAAAGAGTGATTTTGAATAGTGTCCCATGTTGTCTTCTACATTAG 1200
1513  TAATTTCTATTNNAAC--ATGAAATTGCCTTTTATGAGTGATACCCATTTT-ANGGCCTCT 1569
```

```
||||| 1201 TAATTTCTGATTAAACCATGAAATTTGCCCTTATGAGTGATACCCATTGAGGCGCTCT 1260
||| 1570 AANCCCTT---CATTTGACTACGCTGAAGGAGGAAAGCGGAGAGGTAATTTTAT 1626
||| 1261 AANCCCTTCAATTTGTTACTAGCTGAAG-AGGAAAGCGGAAGATGGTAATTTTTTTT 1319
|| 1627 GGTGA 1631
||
|| 1320 ACGGA 1324

RESULT 11
HSA6077
LOCUS Homo sapiens beta3gal-T5 gene. 1197 bp mRNA linear PRI 11-MAY-2000
DEFINITION
ACCESSION AJ006077
VERSION AJ006077.1 GI:7799920
KEYWORDS beta-1,3-galactosyltransferase; beta3gal-T5 gene.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Anado,M., Carneiro,F. and Clausen,H.
TITLE Cloning and expression of two beta-1,3-galactosyltransferases:
beta3gal-T5 and beta3gal-T6
JOURNAL Unpublished
AUTHORS Anado,M.
TITLE Direct Submission
JOURNAL Submitted (11-MAY-1998); Amado M., Department of Oral Diagnostics,
Royal dental School, Norre Alle 20, 2200 Copenhagen, DENMARK
FEATURES
source 1..1197
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..1197
/gene="beta3gal-T5"
1..1197
/codon_start=1
/product="beta-1,3-galactosyltransferase"
/protein_id="CA991546.1"
/db_xref="GI:7799921"
/translation="MSVGRRIKLLILMANVFYIFMEVSKSSQEKNGKEVII
PKKFKISTPPPEAYNREDEKLNROYNPILSLTNOTGEAGRLSNISLNYCEPDLR
VTSVVTGFNNLPDRFDLLYLRCRNYSLIDOPDKCAKPELLLAIKSLIHPFARRQ
AIRSAGQESNAGNQIVRVFLGQTPPEDNHPDLSDMLKPFSEKQDILMNYRDTF
FNLSLKEVLFLRWYSTCPDTEVFVKDDDDVFVNTHHILNLSLSTAKAKOLFIDV
IHNAGPHRDKLKYIPEVYVYSG.YPPYAGGGFLYSGHLALRLYHITDQVHLPIDD
VYTGMLQKLGVLPEKHKGFRPTDIEKNKNICSYVDLMLVHSRKPQEMDIWSOLO
SAHLKC"
BASE COUNT 347 a 274 c 281 g 295 t
ORIGIN

Query Match 69.2%; Score 1181; DB 9; Length-1197;
Best Local Similarity 99.7%; Pred. No. 7.9e-264;
Matches 1194; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 246 ATGAGTGTTCGAGCTCGAAGAAATAA---GTTTGGGTATCTGATGATGCGCAATGTC 302
||| 1 ATGAGTGTTCGAGCTCGAAGAAATAAAGTCTTGTGGGTATCTGATGATGCGCAATGTC 60
|||
QY 303 TCATTTTATTTATTTGGAAGTCTCCAAAGCAGTAGCCCAAGAAAATAATGAAAAGG 362
|||
DB 61 TCATTTTATTTATTTGGAAGTCTCCAAAGCAGTAGCCCAAGAAAATAATGAAAAGG 120
|||
QY 363 GAAGTAATAATACCAAGAGAAGTTCTGGAAGATATCTACCCCTCCCGAGGCATATCTG 422
|||
DB 121 GAAGTAATAATACCAAGAGAAGTTCTGGAAGATATCTACCCCTCCCGAGGCATATCTG 180
|||
```

```
QY 423 AACCGAGACAAAGAAAGCTGAACGGCAGTACAACCCCATCTGAGTAIGCTGACCAAC 482
||| 181 AACCGAGACAAAGAAAGCTGAACGGCAGTACAACCCCATCTGAGCATGCTGACCAAC 240
||| 483 CAGACGGGGAGGGGGAGGCTCTCCATATTAAGCCATCTGAACCTACTGCGAACCCTGAC 542
||| 241 CAGACGGGGAGGGGGAGGCTCTCCAATATAAGCCATCTGAACCTACTGCGAACCCTGAC 300
||| 543 CTGAGGGTCACGTCGGTGTACGGGTTTAAACAACCTTGCAGGACAGATTTAAAGACATT 602
||| 301 CTGAGGGTCACGTCGGTGTACGGGTTTAAACAACCTTGCAGGACAGATTTAAAGACATT 360
||| 603 CTGCTGTATTTGAGATGCGCAATTTATCACTGCTTATAGATCAGCCGGAATGTCGA 662
||| 361 CTGCTGTATTTGAGATGCGCAATTTATCACTGCTTATAGATCAGCCGGAATGTCGA 420
||| 663 AAGAAACCTTTCTTGTGCTGGCGATTAAAGTCCCTCACTCCACATTTTGCAGAGGCAA 722
||| 421 AAGAAACCTTTCTTGTGCTGGCGATTAAAGTCCCTCACTCCACATTTTGCAGAGGCAA 480
||| 723 GCAATCCGGGAATCTCTGGGGCCCAAGAAACACGAGGGAACCAACCGTGTGCGAGTC 782
||| 481 GCAATCCGGGAATCTCTGGGGCCCAAGAAACACGAGGGAACCAACCGTGTGCGAGTC 540
||| 783 TTCTGCTGGGCAGACACCCCGAGAGGACACACCCCGAGCTTTTCAGATATGCTGAAA 842
||| 541 TTCTGCTGGGCAGACACCCCGAGAGGACACACCCCGAGCTTTTCAGATATGCTGAAA 600
||| 843 TTGAGAGTGAGAACCAACCAAGACATCTTATGTGGAAGTACAGAGACACATTTCTTCAAC 902
||| 601 TTGAGAGTGAGAACCAACCAAGACATCTTATGTGGAAGTACAGAGACACATTTCTTCAAC 660
||| 903 TTGCTCTGAAAGAAAGTGTCTTCTCAGTGGGTAAAGTACTTCTGCGCCAGACACTGAG 962
||| 661 TTGCTCTGAAAGAAAGTGTCTTCTCAGTGGGTAAAGTACTTCTGCGCCAGACACTGAG 720
||| 963 TTTGTTTTCAAGGGCGATGACGATGTTTTTGTGAACACCCCATCACATCTGAAATTTG 1022
||| 721 TTTGTTTTCAAGGGCGATGACGATGTTTTTGTGAACACCCCATCACATCTGAAATTTG 780
||| 1023 AATAGTTTATCAAGACCAACCAAGACATCTTCTCATAGGTGATGTATCCCAATGCT 1082
||| 781 AATAGTTTATCAAGACCAACCAAGACATCTTCTCATAGGTGATGTATCCCAATGCT 840
||| 1083 GGACCTCATCGGGTAAGAAGCTGAAGTACTACATCCCAAGAAAGTGTGTACTCTGGCCTC 1142
||| 841 GGACCTCATCGGGTAAGAAGCTGAAGTACTACATCCCAAGAAAGTGTGTACTCTGGCCTC 900
||| 1143 TACCCACCTTATGCAAGGGAGGGGGTTCCTTCTACTCGGCCACTGCGCCCTGAGGCTG 1202
||| 901 TACCCACCTTATGCAAGGGAGGGGGTTCCTTCTACTCGGCCACTGCGCCCTGAGGCTG 960
||| 1203 TACCATATCACTGACGAGTGCCATCTTACCCCATTTGATGAGCTTTATAGTGAATGTC 1262
||| 961 TACCATATCACTGACGAGTGCCATCTTACCCCATTTGATGAGCTTTATAGTGAATGTC 1020
||| 1263 CTTCAAAAACCTCGGCTCGTTCCAGAGAAACAAAGGCTTCAGGACATTTGATATCGAG 1322
||| 1021 CTTCAAAAACCTCGGCTCGTTCCAGAGAAACAAAGGCTTCAGGACATTTGATATCGAG 1080
||| 1323 GAGAAAACAAATAATACATCTGCTCTATGTAGATCTGATGTTAGTACATAGTAAAA 1382
||| 1081 GAGAAAACAAATAATACATCTGCTCTATGTAGATCTGATGTTAGTACATAGTAAAA 1140
||| 1383 CTTCAAGAGATGATTTGTTGTTGCTCAGTGTGAGAGTGTCTCATTTAAATGCTAA 1439
||| 1141 CTTCAAGAGATGATTTGTTGTTGCTCAGTGTGAGAGTGTCTCATTTAAATGCTAA 1197
|||
```

RESULT 12
AY043479
LOCUS
DEFINITION Mus musculus beta-1,3-N-acetylglucosaminyltransferase mRNA.

AY043479 2672 bp mRNA linear ROD 02-SEP-2001

RESULT 13
AL390027/c

LOCUS
DEFINITION
AL390027 Human DNA sequence from clone RP11-368G3 on chromosome X, complete
sequence.
ACCESSION
VERSION AL390027
KEYWORDS
SOURCE HTG.
ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS
TITLE
JOURNAL
Chapman, J.
Direct Submission
Submitted (06-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerequests@sanger.ac.uk
humquerry@sanger.ac.uk clone requests: clonerequests@sanger.ac.uk
On Oct 1, 2000 this sequence version replaced gi:10279715.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
This sequence is the entire insert of clone RP11-368G3 This
sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. RP11-368G3 is from
the library RPCI-11.2 constructed by the group of Pieter de Jong.
For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.

FEATURES
source
Location/Qualifiers
1..200191
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP11-368G3"
/clone_lib="RPCI-11.2"
1..5390
/note="L1MAL repeat: matches 780..6198 of consensus"
repeat_region 5395..7739
/note="L1MB1 repeat: matches 3353..5691 of consensus"
repeat_region 7803..8244
/note="L1MB1 repeat: matches 5706..6165 of consensus"
repeat_region 8439..8794
/note="178 copies 2 mer aa 55% conserved"
repeat_region 8857..8978
/note="61 copies 2 mer 99 61% conserved"
repeat_region 9316..16924
/note="L1MAL repeat: matches 1389..6304 of consensus"
repeat_region 16978..17379
/note="MSTA repeat: matches 1..422 of consensus"
repeat_region 17394..19014
/note="MSTA-internal repeat: matches 1..1636 of consensus"
repeat_region 19015..19402
/note="MSTA repeat: matches 1..426 of consensus"

repeat_region 22130..22169
/note="L2 repeat: matches 2711..2750 of consensus"
repeat_region 22707..24172
/note="L1MB2 repeat: matches 4655..6150 of consensus"
repeat_region 24594..24774
/note="HALL repeat: matches 1535..1722 of consensus"
repeat_region 24881..29302
/note="L1PA16 repeat: matches 1855..6154 of consensus"
repeat_region 29303..29667
/note="THEL1B repeat: matches 1..364 of consensus"
repeat_region 29668..30215
/note="L1PA16 repeat: matches 1323..1855 of consensus"
repeat_region 30216..30336
/note="L1PB2 repeat: matches 6034..6155 of consensus"
repeat_region 30337..30644
/note="AluSg repeat: matches 2..297 of consensus"
repeat_region 30645..30827
/note="L1PB2 repeat: matches 5854..6034 of consensus"
repeat_region 30828..31914
/note="MER11C repeat: matches 1..1071 of consensus"
repeat_region 31915..32879
/note="L1PB2 repeat: matches 4890..5859 of consensus"
repeat_region 31915..31958
/note="L1PA16 repeat: matches 1013..1323 of consensus"
repeat_region 32855..35026
/note="L1PA15-16 repeat: matches 705..1140 of consensus"
repeat_region 35421..35872
/note="HALL repeat: matches 674..1145 of consensus"
repeat_region 35871..35992
/note="HALL repeat: matches 179..298 of consensus"
repeat_region 36101..36215
/note="HALL repeat: matches 27..144 of consensus"
misc_feature 36283..36679
/note="match: GSS: Em:AQ679890"
misc_feature 38239..38719
/note="match: GSS: Em:AQ241053"
repeat_region 39043..39393
/note="L1M4 repeat: matches 5484..5818 of consensus"
repeat_region 40247..40456
/note="L1MA5 repeat: matches 6083..6293 of consensus"
repeat_region 41291..41600
/note="AluSp repeat: matches 3..311 of consensus"
repeat_region 41678..41810
/note="MIR repeat: matches 76..201 of consensus"
repeat_region 42317..42535
/note="MLT1G repeat: matches 301..547 of consensus"
repeat_region 42611..42777
/note="MLT1G repeat: matches 106..271 of consensus"
repeat_region 43051..48834
/note="L1PA5 repeat: matches 361..6143 of consensus"
repeat_region 48971..49161
/note="MIR repeat: matches 11..208 of consensus"
misc_feature 51713..51814
/note="match: GSS: Em:AQ486704"
repeat_region 51825..52294
/note="L1PA7 repeat: matches 5676..6145 of consensus"
repeat_region 52519..53592
/note="L1M1 repeat: matches 1384..241 of consensus"
repeat_region 53593..53889
/note="AluY repeat: matches 1..298 of consensus"
repeat_region 53890..54000
/note="L1M1 repeat: matches 241..126 of consensus"
repeat_region 54010..54781
/note="L1PA3 repeat: matches 5376..6146 of consensus"
repeat_region 54801..54858
/note="L1M1 repeat: matches 5385..5437 of consensus"
repeat_region 54871..56799
/note="L1M1 repeat: matches 130..1815 of consensus"
repeat_region 56815..57963
/note="L1PA13 repeat: matches 651..606 of consensus"
repeat_region 57986..58552
/note="L1PA13 repeat: matches 758..1326 of consensus"
repeat_region 58510..58615

```
/note="L1 repeat: matches 2444. .2549 of consensus"
58612. .59898
/note="L1MA2 repeat: matches 4985. .6290 of consensus"
60790. .60855
/note="L2 repeat: matches 2644. .2709 of consensus"
60987. .61271
/note="AluX repeat: matches 1. .283 of consensus"
61004. .61457
/note="match: S7S: Em:HSC71D4"
61362. .61583
/note="L1MC/D repeat: matches 5422. .5651 of consensus"
61690. .61749
/note="30 copies 2 mer ca 75% conserved"
63509. .63631
/note="AluYa5/8 repeat: matches 187. .309 of consensus"
63758. .63819
/note="31 copies 2 mer ca 83% conserved"
64852. .64887
/note="18 copies 2 mer tt 83% conserved"
64901. .68363
/note="L1PB1 repeat: matches 2691. .6151 of consensus"
68363. .70591
/note="L1PA13 repeat: matches 3895. .6156 of consensus"
70603. .71745
/note="L1M1 repeat: matches 1803. .2941 of consensus"
71727. .73819
/note="L1PBa repeat: matches -279. .2760 of consensus"
73807. .74870
/note="L1PBa repeat: matches -1537. .412 of consensus"
75305. .75479
/note="L1M2 repeat: matches 307. .478 of consensus"
75480. .75832
/note="THE1B repeat: matches 1. .364 of consensus"
75833. .76922
/note="L1M2 repeat: matches 478. .1582 of consensus"
76975. .77185
/note="L1M2 repeat: matches 1773. .1587 of consensus"
77293. .77665
/note="MLT:A1 repeat: matches 3. .365 of consensus"
77745. .77964
/note="L1 repeat: matches 1880. .2112 of consensus"
78016. .78085
/note="L1 repeat: matches 2203. .2272 of consensus"
78087. .80520
/note="L1PA5 repeat: matches 3693. .6141 of consensus"
80541. .81238
/note="L1 repeat: matches 2254. .3032 of consensus"
81231. .81559
/note="L1ME1 repeat: matches 5763. .6101 of consensus"
81653. .81904
/note="MIR repeat: matches 7. .262 of consensus"
83147. .83435
/note="AluY repeat: matches 2. .299 of consensus"
83506. .84072
/note="MER7C repeat: matches 38. .602 of consensus"
85050. .85188
/note="L1MA3 repeat: matches 6166. .6304 of consensus"

Query Match 55.9%; Score 954.2; DB 9; Length 200191;
Best Local Similarity 81.9%; Pred. No. 1.4e-210;
Matches 1205; Conservative 7; Mismatches 210; Indels 49; Gaps 8;

QY 98 CCGCTCCGGCGCGCATGGAGCTGAGCTGCGGGTGGCGGCTCAGCGCGG 157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125618 CCCACCCCGCGCGCCCATAGAGCTGAGCTGTGGCTTCTCAGGCTCAGCGGTG 125559

QY 158 GAGCGCGCGG---ACGTGGATGTGGCGGCACTCTCC-----GCCCTGCC 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125558 GAGCGGTGGGAAAGTGGTATGGCAGCGCATTCACACCCCGCCGCCACC 125499

QY 202 CCGCGCGCGGAGCTGGAGCTGTCTCCGGACACATATGAAATAGTGTGGAGCTC 261
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125498 CCCGACCCGAGCTGAGTCTTGGACAGATATGACATGAGTCTTCAAGCTT 125439
```

```
QY 262 GAACAATAAAGTTGTTGGGTATCCTGATGATGCGAATGCTTCAITATTATTATG 321
DB 125438 GAAGAATAAAGTTGTTGGGTAACTGATGATGCGAATGCTTCAITATTATTATG 125379

QY 322 AAGTCTCCAAAGCAGTAGCAAGAAAAAATGAAAA----- 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125378 AAGCTTCCAAAGCAGTAGCCAATATAATAATAATAATAATAATAATAATAATG 125319

QY 360 -----GGGAAGTAATAATACCCAAAGAGAGTTCTTGGAGATATACCTCCCGAGG 414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125318 AAGGTGGTTTAATAATAATACCCAAAGAGAGTTTGGAGAGATTGCTACACCTCTCGAGG 125259

QY 415 CATATGGAACCGAGACGAGAGAGAGCTGAACCGGAGTACACACCCATCTCAGCATGC 474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125258 CACATGGAACGAGAGACAAAGAAAGCTGAACAAGTGGTACAAATCCATCTCAGTAGT 125199

QY 475 TGACCAACAGAGCGGGGAGCGGAGGCTCTCCAATATAAAGCCATCTCAACTACTGCG 534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125198 TGGCCACACAGC- GGGGAAGTGTACGGGTTTTTCAGTATAAGCCATCTGAATTACTGTG 125140

QY 535 AACTGACCTGAGGGTACCTCGGTGGTTACGGGTTTTAAACAACITGCGCGACAGATTTA 594
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125139 AAGCTGACTTGAGGTTCACTGACGTGGTTACAGATTTTAAACAATTTCTCTGGACAGATGTG 125080

QY 595 AAGACTTTCTGCTGTTTTCAGATGCGGCAATATTACTGCTTATAGATCAGCCGATA 654
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125079 ATGACTTTCTGATGATTTGAGATGCTGCAATGATTCACCTGCTTTAGATCAGCCGATA 125020

QY 655 AGTGTGCAAGAAACCTTCTTGTGCTGGCGAATTAAGTCCCTCACTCCACATTTTGCCA 714
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125019 AGTGTGCAAGAAAGCACITCTTATTCTGGTGATTAAGTACCCTCACTCCCATTTTGCCA 124960

QY 715 GAAGGCAAGCAATCCGGGAATCTCGGGCCCAAGAAAGCAACGAGGAAACCAACGGTGG 774
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 124959 CAAGGCAAGTGATTCGGGATTCCTGGGTGAGAGAAACCAACGTTGGAAAGGTTGA 124900

QY 775 TGGAGCTTCTGCTGCTGGGCGAGACACCCCGAGGAGCAACCCCGACCTTTCAGATA 834
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 124899 TGAAGATTTCTGCTGGGTGATACCCCGAGGTTCAACCCCGGAGTTTCTTGATA 124840

QY 835 TGTGAAATTTGAGATGAGAGCA-CCAAAGACATTTATGTGGAACACTACAGACACT 893
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 124839 TGTGAAATTTGAGATGAGAGCAACCCCGAGATAGTTTTATCGGAGCTATAGACACT 124780

QY 894 TCTTCAACTGTCTCTGAGAGAGTGTCTTTCTCAGGTGGGTAACTCTCTCGGCCA 953
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 124779 TTCTTCAAGTGTCTCTGAAAGAAGTGTGTTTCTC- GCTGGGTAGTACTTCTGCCCCA 124721

QY 954 GACACTGAGTTTGTTCAGGGCGATGACGATGTTTTTGTGAACCCCATCATCCTG 1013
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 124720 GACACTGAGTTGTTTTCAGGGCGAAGATGATGTTTTGTGAACCCCATCATCCTG 124661

QY 1014 AATTACTTGAATAGTTTATCCAAG- ACCAAAGCAAGATCTCTTCATAGGTGATGTAT 1072
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 124660 AATTACTTGAATAGTTTATCCAAGAAAAAAGCAAGATCTGTTTATAGTGTATGTAT 124601

QY 1073 CCACATGCTGGACCTCATCGGATAGAGCTGAAGTACTACTATCCAGAAAGTTGTTTA 1132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 124600 CCACATGCTGGACCTCATCGGATAGAGCTGAATACTACTATCTAGAAGTTGTTTA 124541

QY 1133 CTCTGGCTCTACCCACCTTATGCA- GGGGAGGGGGTTCCTCTACTCCGCCACCTG 1190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 124540 CTCTGTCTCTACTCACCTTATGCAAGGGGAGGGGGATTCCTCTAGTCCAGCCACCTG 124481

QY 1191 GCCTGAGGCTGTACCATATCACTGACAGGTCCATCTCTACCCCATTTGATGACGTTAT 1250
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 124480 GCTCTGAGGCTGTACCATATCACTGACTGGTCCATCTCTAACTTATGATGATTTTAT 124421

QY 1251 ACTGGAATGTGCTTCCAGAAACTCGGCTCTGTTCCAGAGAAACACAAAGGCTTCAGGACA 1310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 124420 ACTGCAATGTGCCTTCAAAAACCTTAGCTTGTTCAGAGAAGCAAAAGGCTTCAGGACA 124361
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2003, 12:46:17 ; Search time 482.17 Seconds
(without alignments)
9556.669 Million cell updates/sec

Title: US-09-804-357b-13
Perfect score: 1707
Sequence: 1 acgcgtccgcagcggcag.....aatctgttgatggccctt 1707

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03.*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1694.8	99.3	1707	21	AA12716
2	1694.8	99.3	1707	23	AA167869
3	1694.8	99.3	1707	24	AA167869
4	1531.4	89.7	1831	25	ABV75082
5	1504.2	88.1	1912	22	AAF29255
6	1488	87.2	2745	19	AAV66367
7	1231.4	72.1	2442	22	AAH14972
8	1229.8	72.0	2536	21	AAZ87186

9	1038.4	60.8	1194	21	AAZ88487	Human brainiac pro
10	1002.2	58.7	2676	21	AAZ88486	Murine brainiac en
11	1002.2	58.7	2676	21	AAZ87185	Murine brainiac cd
12	931.4	54.6	1260	25	ABV75081	Murine beta3Gnt ge
13	906.4	53.1	1196	21	AA12714	cDNA encoding murf
14	906.4	53.1	1196	23	AA167865	Murine LIG46 poly
15	906.4	53.1	1196	24	AA167866	Murine LIG46 cDNA
16	906.4	53.1	1194	23	AA167866	Murine LIG46 poly
17	717.2	42.0	832	22	AAH07800	Human cDNA clone (
18	533.6	31.3	556	20	AAV89848	EST clone CW1306.
19	431.4	25.3	857	19	AAV66384	cDNA clone W26453.
20	328.2	19.2	480	19	AAV66390	cDNA clone H47991.
21	284.8	16.7	303	19	AAV66395	cDNA clone H00589.
22	279	16.3	282	19	AAV66398	cDNA clone H00116.
23	279	16.3	282	19	AAV66375	cDNA clone HTA441
24	267.4	15.7	428	19	AAV66400	cDNA clone AA38183
25	264	15.5	266	19	AAV66401	cDNA clone AA37708
26	264	15.5	266	19	AAV66376	cDNA clone HTABE60
27	257	15.1	259	19	AAV66402	cDNA clone H13126
28	257	15.1	259	19	AAV66378	cDNA clone H518A68
29	223.4	13.1	473	21	AAZ88389	Human secreted pro
30	203.4	11.9	1245	24	ABK89794	Human galactosyltr
31	175.8	10.3	368	19	AAV66404	cDNA clone AA37708
32	174.4	10.2	1134	21	AA58791	DNA encoding the b
33	174.4	10.2	1134	24	AB553140	Degenerate DNA enc
34	174.4	10.2	1134	24	ABD39776	Human zsssp6 degen
35	167.8	9.8	1360	24	AAZ35225	Human TRNFR-10 cDN
36	167.6	9.8	1336	24	ABX97158	Human NOV100a cDNA
37	167.6	9.8	1359	23	AA584438	DNA encoding novel
38	166.2	9.7	3356	25	ABX63848	Human cDNA #848 d1
39	164.8	9.7	2003	24	ABK89793	cDNA encoding huma
40	163.2	9.6	1116	19	AAV49599	Human epidermold c
41	163.2	9.6	1208	24	AA516945	Human beta1.3-N-ac
42	163.2	9.6	2186	19	AAV49598	Human epidermold c
43	163.2	9.6	2186	24	ABL41994	Nucleotide sequenc
44	163.2	9.6	2186	25	ABZ78132	Human cancer-relat
45	163.2	9.6	2186	25	ABV76854	Nucleotide sequenc

ALIGNMENTS

RESULT 1
AA12716
ID AA12716 standard; cDNA; 1707 BP.
XX
AC AA12716;
XX 25-JUL-2000 (first entry)
XX cDNA encoding human LIG46 polypeptide.
DE
XX
XX Leptin: LIG46; body weight; leptin inducible gene: obesity; cachexia;
XX LIG56; tgtp; LRG-47; KC10-II; Stral3; ss.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 245..1439
XX FT /*tag= a
XX FT /product= *LIG46*
XX
XX WO200015826-A2.
XX
XX PD 23-MAR-2000.
XX
XX PF 10-SEP-1999; 99WO-US20722.
XX
XX PR 10-SEP-1998; 98US-0150857.
XX PR 29-OCT-1998; 98US-0106378.
XX PR 19-NOV-1998; 98US-0195896.
XX PR 15-APR-1999; 99US-0292228.
XX

Db 1621 NTTATGGTGAATGGCAGGATATTGGTCTGACTTACCGNTAGGGGANTTTAAACATGGNC 1680

Qy 1681 CTTTTGAATCTGTTGGATGGCCCTT 1707

Db 1681 CTTTTGAATCTGTTGGATGGCCCTT 1707

RESULT 2

AA167869

ID AA167869 standard: cDNA: 1707 BP.

XX AA167869;

DT 13-MAR-2002 (first entry)

DE Human LIG46 polypeptide encoding cDNA.

XX LIG46; LIG56; leptin induced gene; Tgtp: LRG-47; RC10-II; Stral3;
KW anorectic; anabolic; antisense therapy; human; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
FT CDS 246..1439
FT /tag= a
FT /product= "LIG46"

XX US2001024808-A1.

XX 27-SEP-2001.

XX 12-MAR-2001; 2001US-0804357.

XX 29-OCT-1998; 98US-106378P.

XX 19-NOV-1998; 98US-0195895.

XX 10-SEP-1998; 98US-0150857.

XX (MILL-) MILLENNIUM PHARM INC.

XX White D, Zhou J, Tartaglia LA;

XX WPI; 2001-624963/72.

XX P-PSDB; AAG66118.

XX An isolated polypeptide useful for treating a weight disorder and for
PT screening compounds that may inhibit protein activity comprises a
PT leptin-induced protein -

XX Example 2; Fig 7; 46pp; English.

XX The invention relates to genes whose expression are induced by leptin.
CC LIG46 and LIG56 are novel leptin induced genes (LIG), while four other
CC genes such as Tgtp, LRG-47, RC10-II and Stral3 have been previously
CC identified. The leptin induced proteins can be expressed by standard
CC recombinant methodology. The proteins and encoding polynucleotides may
CC be used in screening assays to identify compounds that may bind to it.
CC Administering a molecule, e.g., an antisense molecule, which reduces
CC expression of activity of protein selected from LIG46, LIG56, Tgtp,
CC LRG-47, RC10-II, and Stral3 may be used to treat a weight disorder. The
CC method may further comprise administering leptin. The present sequence
CC represents a cDNA encoding a human LIG46 protein.

XX Sequence 1707 BP; 452 A; 403 C; 435 G; 400 T; 17 other;

SQ Query Match 99.3%; Score 1694.8; DB 23; Length 1707;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1707; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACGGTCGCGCAGCGGCGGAGCGGAGCGGCAACAAAGTCCGGAGGCTAGCAGAGCCA 60
|||||

Db 1 ACGGTCGCGCAGCGGCGGAGCGGAGCGGCAACAAAGTCCGGAGGCTAGCAGAGCCA 60
|||||

Qy 61 ACGCGGACAGTCCCTGCGCGCGACACCGCGGCGGCCCGCCCTCCGGGCGCGGCATGG 120

Db 61 ACGCGGACAGTCCCTGCGCGCGACACCGCGGCGCGCGTCCGGGCGCGCGCATGG 120
|||||

Qy 121 ACGGTGAGTGC CGCGGTGCGCGGTGAGCCCGCGGAGCGCGCGGAGTGATGTGG 180
|||||

Db 121 ACGGTGAGTGC CGCGGTGCGCGGTGAGCCCGCGGAGCGCGCGGAGTGATGTGG 180
|||||

Qy 181 CCGGATCTCCCGCCCTTGCCTCCCGCCCGCGAGCTGGAGCTGCTCCCGGACAAGATAT 240
|||||

Db 181 CCGGATCTCCCGCCCTTGCCTCCCGCCCGCGAGCTGGAGCTGCTCCCGGACAAGATAT 240
|||||

Qy 241 GAGAAATGAGTGTGGACGTCGAAAGATAAAGTTGTTGGGTATCTCTGATGCGCAATG 300
|||||

Db 241 GAGAAATGAGTGTGGACGTCGAAAGATAAAGTTGTTGGGTATCTCTGATGCGCAATG 300
|||||

Qy 301 TCTTCATTTATTTATTTATGGAAGTCTCCAAAAGCAGTAGCCCAAGAAAAAATGGAAG 360
|||||

Db 301 TCTTCATTTATTTATTTATGGAAGTCTCCAAAAGCAGTAGCCCAAGAAAAAATGGAAG 360
|||||

Qy 361 GGAAGTATAATACCCARAGAGAGTCTCTGGAAGATATCTACCCCTCCCGGAGCATACT 420
|||||

Db 361 GGAAGTATAATACCCARAGAGAGTCTCTGGAAGATATCTACCCCTCCCGGAGCATACT 420
|||||

Qy 421 GGAACCGAGAGCAAGAGAGTGAACCGGCGAGTACAACCCCATCTCTGAGCATGCTGACCA 480
|||||

Db 421 GGAACCGAGAGCAAGAGAGTGAACCGGCGAGTACAACCCCATCTCTGAGCATGCTGACCA 480
|||||

Qy 481 ACCAGACGGGGAGCGGCGGAGGCTCTCCAATATAAGCCATCTGAACACTAGTGGAACTG 540
|||||

Db 481 ACCAGACGGGGAGCGGCGGAGGCTCTCCAATATAAGCCATCTGAACACTAGTGGAACTG 540
|||||

Qy 541 ACTGAGGTCACGTCGCTGCTTACGGGTTTAAACACTTTGCCGAGACAGATTTAAAGACT 600
|||||

Db 541 ACTGAGGTCACGTCGCTGCTTACGGGTTTAAACACTTTGCCGAGACAGATTTAAAGACT 600
|||||

Qy 601 TTCTGCTGTTATTTGAGATGCGCAATTAATCACTGCTTATAGATCAGCCGAGTAAGTGTG 660
|||||

Db 601 TTCTGCTGTTATTTGAGATGCGCAATTAATCACTGCTTATAGATCAGCCGAGTAAGTGTG 660
|||||

Qy 661 CAAAGAAACCTTTCTTCTGCTGGCGATTAACTCCCTCACTCCACATTTTGCCAGAAGGC 720
|||||

Db 661 CAAAGAAACCTTTCTTCTGCTGGCGATTAACTCCCTCACTCCACATTTTGCCAGAAGGC 720
|||||

Qy 721 AAGCAATCCGGGAATCTCTGGGCGCAAGAAACAGCAGGAGCAACCAACCGTGGTGGAG 780
|||||

Db 721 AAGCAATCCGGGAATCTCTGGGCGCAAGAAACAGCAGGAGCAACCAACCGTGGTGGAG 780
|||||

Qy 781 TCTTCTGCTGGCGCACACACCCCGAGGACACCAACCCCGACCTTTTCAGATATGCTGA 840
|||||

Db 781 TCTTCTGCTGGCGCACACACCCCGAGGACACCAACCCCGACCTTTTCAGATATGCTGA 840
|||||

Qy 841 AATTTGAGTGAAGACCAAGACATCTTATGTGGAACCTACAGAGACACTTTCTTCA 900
|||||

Db 841 AATTTGAGTGAAGACCAAGACATCTTATGTGGAACCTACAGAGACACTTTCTTCA 900
|||||

Qy 901 ACTTGCTCTGAGGAGTCTGTTCTTCAGTGGGTAACTACTCTCTGCCAGACACTG 960
|||||

Db 901 ACTTGCTCTGAGGAGTCTGTTCTTCAGTGGGTAACTACTCTCTGCCAGACACTG 960
|||||

Qy 961 AGTTTGTGTTTCAAGGGCGATGAGCATGTTTTTGTGAAACCCCATCAGATCCTGAATTA 1020
|||||

Db 961 AGTTTGTGTTTCAAGGGCGATGAGCATGTTTTTGTGAAACCCCATCAGATCCTGAATTA 1020
|||||

Qy 1021 TGAATAGTTTATCCAGACCAAGCCAAAGATCTCTTCATAGGTGATGATCCACAATG 1080
|||||

Db 1021 TGAATAGTTTATCCAGACCAAGCCAAAGATCTCTTCATAGGTGATGATCCACAATG 1080
|||||

Qy 1081 CTGGACCTCATCGGGATAAGAGCTGAAGTACTACATCCAGAGTGTGTTTACTCTGGCC 1140
|||||

Db 1081 CTGGACCTCATCGGGATAAGAGCTGAAGTACTACATCCAGAGTGTGTTTACTCTGGCC 1140
|||||

Qy 1141 TCTACCCACCTTATGACGGGGAGGGGGTTCCTCTACTCCGCCACCTGGCCCTGAGGC 1200
|||||

```

Db 1141 TCTACCCACCTATGACGGGGAGGGGGTTCCTCTACTCCGGCCACCTGGCCCTGAGGC 1200
Qy 1201 TGACCATATCACTGACCGAGTCCATCTCTACCCCATTCATGACGTTATATAGTGAATGT 1260
Db 1201 TGTACCATATCACTGACCGAGTCCATCTCTACCCCATTCATGACGTTTATATAGTGAATGT 1260
Qy 1261 GCCTTCAGAAACTCGGCCTCGTTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATCG 1320
Db 1261 GCCTTCAGAAACTCGGCCTCGTTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATCG 1320
Qy 1321 AGAGAAAAACAATAAATACATCTGCTCCTATGATATCTGATCTGATCTGATCTGATCTGAT 1380
Db 1321 AGAGAAAAACAATAAATACATCTGCTCCTATGATATCTGATCTGATCTGATCTGATCTGAT 1380
Qy 1381 AACCTCAAGAGATGATGATATTTGGTCTCAGTTCGAGAGTCTCATTTAAATGCTAAA 1440
Db 1381 AACCTCAAGAGATGATGATATTTGGTCTCAGTTCGAGAGTCTCATTTAAATGCTAAA 1440
Qy 1441 ATAGATACAAACTCAATTTKGSATWGRAAGGGGTWTTTGRATWGYCCCATGTTGGGGT 1500
Db 1441 ATAGATACAAACTCAATTTKGSATWGRAAGGGGTWTTTGRATWGYCCCATGTTGGGGT 1500
Qy 1501 CTCACATAGAGTAATTTCTATTNAANCATGAATGGCTTTATCAGTGATACCATTT 1560
Db 1501 CTCACATAGAGTAATTTCTATTNAANCATGAATGGCTTTATCAGTGATACCATTT 1560
Qy 1561 ANGCGCTCTAANCCTTCATTGTCAGTGAAGAGGAAAGCGGAGAGTAATTT 1620
Db 1561 ANGCGCTCTAANCCTTCATTGTCAGTGAAGAGGAAAGCGGAGAGTAATTT 1620
Qy 1621 NTTTATGGTGAATGGCAGGATATTTGGTCTGACTTACCGNTAGGGGANTTTAAACTGGNC 1680
Db 1621 NTTTATGGTGAATGGCAGGATATTTGGTCTGACTTACCGNTAGGGGANTTTAAACTGGNC 1680
Qy 1681 CTTTTTGAATCTGTTTGGATGGCCCTT 1707
Db 1681 CTTTTTGAATCTGTTTGGATGGCCCTT 1707

RESULT 3
AAD46673
ID AAD46673 standard; cDNA; 1707 BP.
AC AAD46673;
XX
DT 27-JAN-2003 (first entry)
XX
DE Human LIG46 cDNA.
XX
KW LIG46; leptin; low body weight; chromosomal mapping; tissue typing;
KW forensic biology; transgenic; gene therapy; antianorectic; human;
KW chromosome 2; gene; ss.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 246..1439
FT /*tag= a
FT /product= "Human LIG46 protein"
FT sig_peptide 246..341
FT /*tag= b
FT mat_peptide 342..1436
FT /*tag= c
FT /product= "Human mature LIG46 protein"
XX
XX WO200274905-A2.
PN
XX
XX 26-SEP-2002.
PD
XX
XX 20-NOV-2001; 2001WO-US43345.
PF
XX
XX 21-NOV-2000; 2000US-0717778.
PR
XX
XX

```

```

PA (MILL-) MILLENNIUM PHARM INC.
XX
P2 White DW, Zhou J, Tartaglia LA, Stricker-krongrad A, Clausen H;
XX
XX WPI; 2002-759886/B2.
DR P-PSDB: AAE29093.
XX
XX New isolated nucleic acid molecules, designated as leptin-induced genes
PT 46 (LIG46), useful for treating a subject having a disorder
PT characterized by undesirable level of LIG46 expression or activity,
PT such as low body weight
XX
XX Claim 15: Fig 5; 90pp: English.
XX
XX The invention relates to LIG46, a gene whose expression is induced by
CC leptin. LIG46 DNA and protein are useful in treating a subject having
CC a disorder characterised by undesirable level of LIG46 expression or
CC activity, such as low body weight. They are also useful in a screening
CC assay, chromosomal mapping, tissue typing and forensic biology. The
CC probes based on the LIG46 nucleotide sequence are useful for detecting
CC transcripts or genomic sequences encoding the same or related proteins.
CC The LIG46 polypeptides are useful as immunogens for raising anti-LIG46
CC antibodies. The host cells are useful for producing non-human transgenic
CC animals. LIG46 DNA is used in gene therapy. The present sequence is
CC human LIG46 cDNA. LIG46 gene is located at chromosome 2.
XX
XX Sequence 1707 BP; 452 A; 403 C; 435 G; 400 T; 17 other;
XX
Query Match 99.38; Score 1694.8; DB 24; Length 1707;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1707; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ACGGGTCCGCGCAGCGGCGAGCGGCGGCAACAAGTCCCGAGGCTAGCAGACCCA 60
Db 1 ACGGGTCCGCGCAGCGGCGAGCGGCGGCAACAAGTCCCGAGGCTAGCAGACCCA 60
Qy 61 AGCGGAGCAGTCCCTGCGGCCGACACCGCGGCGCGCGTCCGGGGCGCGGCATGG 120
Db 61 AGCGGAGCAGTCCCTGCGGCCGACACCGCGGCGCGCGTCCGGGGCGCGGCATGG 120
Qy 121 AGCGTGAGCTCGCGGCTCGCGGCGGCTGAGCGCGGCGGAGCGGCGGAGTGTGG 180
Db 121 AGCGTGAGCTCGCGGCTCGCGGCGGCTGAGCGCGGCGGAGCGGCGGAGTGTGG 180
Qy 181 CGCGCATCTCCCGCCCTTGCCCGCGCGCGCGCGCGGAGCTGCTCCCGACAAATAT 240
Db 181 CGCGCATCTCCCGCCCTTGCCCGCGCGCGCGCGGAGCTGCTCCCGACAAATAT 240
Qy 241 GAGAAATCAGTGTGGAGCTCGAAGATTAAGTTGTTGGGTATCTCTGATGTGCAATG 300
Db 241 GAGAAATCAGTGTGGAGCTCGAAGATTAAGTTGTTGGGTATCTCTGATGTGCAATG 300
Qy 301 TCTTCATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 360
Db 301 TCTTCATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 360
Qy 361 GGGAAAGTAAATATACCCAAAGAGAGTTCGGAAGATATCTACCCCTCCCGAGCATPACT 420
Db 361 GGGAAAGTAAATATACCCAAAGAGAGTTCGGAAGATATCTACCCCTCCCGAGCATPACT 420
Qy 421 GGAACCGAGAGCAAGAGAGCTGAACCGGCGAGTGAACCGGCGAGTGAACCGGCGAGTGAAC 480
Db 421 GGAACCGAGAGCAAGAGAGCTGAACCGGCGAGTGAACCGGCGAGTGAACCGGCGAGTGAAC 480
Qy 481 ACCAGACGGGGAGGCGGCGAGGCTCTCCAATATAAGCCATCTGAACCTACTGCGAAGCTG 540
Db 481 ACCAGACGGGGAGGCGGCGAGGCTCTCCAATATAAGCCATCTGAACCTACTGCGAAGCTG 540
Qy 541 ACCTGAGGTCACGTCGGTGGTTACGGGTTTTTAAACAACCTTCCCGGACAGATTTAAAGACT 600
Db 541 ACCTGAGGTCACGTCGGTGGTTACGGGTTTTTAAACAACCTTCCCGGACAGATTTAAAGACT 600
Qy 601 TTCTGCTGATTTGAGATGCCGCAATTTATTCACCTGCTTATAGATCAGCCGGATAAGTGTG 660

```


Db	1681	CTTTTGAATCTGTTGGATGCCCTT	1707
RESULT 4			
ABV75082			
ID	ABV75082	standard; DNA; 1831 BP.	
XX	AC	ABV75082;	
XX	DI	19-FEB-2003 (first entry)	
XX	XX	Human beta3Gnt gene polynucleotide sequence.	
XX	DE	Beta3Gnt; transgenic; beta-1,3-N-acetylglucosaminyltransferase;	
KW	KW	disease system; gene; human; ds.	
XX	XX	Homo sapiens.	
XX	XX	Key Location/Qualifiers	
PH	PH	235, 1429	
FT	FT	/*tag= a	
FT	FT	/product= "beta3Gnt"	
XX	XX	W0200279413-A2.	
XX	XX	10-OCT-2002.	
XX	XX	29-MAR-2002; 2002WO-US09645.	
XX	XX	29-MAR-2001; 2001US-280706P.	
PR	PR	28-MAR-2002; 2002US-0280706.	
XX	XX	(DELT-) DELTAGEN INC.	
XX	XX	Leviten MW, Phillips R;	
XX	XX	WP.: 2003-067437/06.	
DR	DR	P-PSDB; ABB82649.	
XX	XX	New transgenic mouse comprising a disruption in a beta3Gnt gene, as in	
PT	PT	vivo model to study various disease states or conditions in which	
PT	PT	beta3Gnt may be implicated or involved, such as abnormal cell growth,	
PT	PT	cancer and metastasis	
XX	XX	Disclosure: Fig 3: 55pp: English.	
PS	PS	The invention relates to a transgenic mouse comprising a disruption in a	
CC	CC	beta3Gnt (beta-1,3-N-acetylglucosaminyltransferase) gene, where there is	
CC	CC	no native expression of beta3Gnt gene. The transgenic mice may be used	
CC	CC	as in vivo model to study various disease states or conditions in which	
CC	CC	beta3Gnt may be implicated or involved, such as abnormal cell growth,	
CC	CC	cancer and metastasis, and to evaluate various treatments or to identify	
CC	CC	agents for treating disease states or conditions, such as anxiety or	
CC	CC	depression. Animal-based disease systems may be used to identify	
CC	CC	compounds capable of ameliorating disease symptoms, as test substrates	
CC	CC	for the identification of drugs, pharmaceuticals, therapies and	
CC	CC	interventions that may be effective in treating a disease or other	
CC	CC	phenotypic characteristic of the animal. The present sequence represents	
CC	CC	a human beta3Gnt polynucleotide sequence.	
XX	XX	Sequence 1831 BP; 497 A; 423 C; 443 G; 468 T; 0 other;	
SQ	SQ		
Query Match 89.78; Score 1531.4; DB 25; Length 1831;			
Best Local Similarity 97.5%; Pred. No. 0;			
Matches 1585; Conservative 8; Mismatches 25; Indels 7; Gaps 4;			
OY	13	AGCGGCGAGCGGACGACCGGCAACAGTGC CGCGAGGCTAGCAGAGCCAGCCGAGCAGT	72
Db	3	AGCGGCGAGCGGACGACCGGCAACAGTGC CGCGAGGCTAGCAGAGCCAGCCGAGCAGT	62
QY	73	CCCTGCCCGCGACACCGCCGCGCCGCGCCGCTCCCGGGGCGCCGCGCATGGAGCGTGAGCTGC	132
Db	63	CCCTGCCCGCGACACCGCGGCGCGCCGCTCCCGGGGCGCCGCGCATGGAGCGTGAGCTGC	122

```
QY 133 GCGGCTCGCGGGCTGAGCGCGGAGCGCGCGGAGCTGGATGCTGCGCGGATCTCCC 192
Db 133 GCGGCTCGCGGGCTGAGCGCGGAGCGCGCGGAGCTGGATGCTGCGCGGATCTCCC 182
QY 193 GCGCTTGGCCCCCGCCCCCGAGCTGAGCTGCTCCCGGACAAGATATGAGAAATGAGTG 252
Db 183 GCGCTTGGCCCCCGCCCCCGAGCTGAGCTGCTCCCGGACAAGATATGAGAAATGAGTG 242
QY 253 TTGGAGCTCGAAGATAAAGTTGTTGGGTATCCTGATGATGCGCAAAATGCTTCAATTAT 312
Db 243 TTGGAGCTCGAAGATAAAGTTGTTGGGTATCCTGATGATGCGCAAAATGCTTCAATTAT 302
QY 313 TTATTATGAACTCTCCAAAAGCAGTAGCCAAGAAAAAATGGAAGGGGAAGTAATAA 372
Db 303 TTATTATGGAAGCTTCCAAAAGCAGTAGCCAAGAAAAAATGGAAGGGGAAGTAATAA 362
QY 373 TACCCAAAGAGAAAGTTCTGGAGATATCTACCCCTCCCGAGGCATCTGGAACCGAGAGC 432
Db 363 TACCCAAAGAGAAAGTTCTGGAGATATCTACCCCTCCCGAGGCATCTGGAACCGAGAGC 422
QY 433 AAGAGAGCTGAACGGGAGTACACCCCATCCTGAGCATGCTGACCAACAGAGCGGG 492
Db 423 AAGAGAGCTGAACGGGAGTACACCCCATCCTGAGCATGCTGACCAACAGAGCGGG 482
QY 493 AGGCGGGAGGCTCTCCCAATATAAGCCATCTGAACCTACTGCGAAGCTGAGGGTCA 552
Db 483 AGGCGGGAGGCTCTCCCAATATAAGCCATCTGAACCTACTGCGAAGCTGAGGGTCA 542
QY 553 CGTCGGTGGTTACGGGTTTAACTACTGCGGAGCAGATTTAAGACTTTTCTGCTGTAIT 612
Db 543 CGTCGGTGGTTACGGGTTTAACTACTGCGGAGCAGATTTAAGACTTTTCTGCTGTAIT 602
QY 613 TGAGATGCGGCAATTAATCACTGCTTATAGATCAACCGGATAGCTGCAAGAAACCT 672
Db 603 TGAGATGCGGCAATTAATCACTGCTTATAGATCAACCGGATAGCTGCAAGAAACCT 662
QY 673 TCTTGTGCTGCGGATTAAGTCCCTCACTCCACATTTTCCAGAGGCAAGCAATCCGGG 732
Db 663 TCTTGTGCTGCGGATTAAGTCCCTCACTCCACATTTTCCAGAGGCAAGCAATCCGGG 722
QY 733 AATCTGGGGCAAGAAAGCAACGAGGAAACCAACCGTGGTGGAGCTCTTCTGCTGG 792
Db 723 AATCTGGGGCAAGAAAGCAACGAGGAAACCAACCGTGGTGGAGCTCTTCTGCTGG 782
QY 793 GCCAGACACCCCGAGGACACACCCCGACCTTTTCAGATATGCTGAAATTTGAGAGTG 852
Db 783 GCCAGACACCCCGAGGACACACCCCGACCTTTTCAGATATGCTGAAATTTGAGAGTG 842
QY 853 AGAAGCACCAAGACATTTCTATGTGGAATACAGAGACACTTTCTTCAACTTGTCTGTA 912
Db 843 AGAAGCACCAAGACATTTCTATGTGGAATACAGAGACACTTTCTTCAACTTGTCTGTA 902
QY 913 AGGAAGTCTGTTCTCAGGTGGTAACTACTTCTGCGGAGACACTGAGTTGTCTCA 972
Db 903 AGGAAGTCTGTTCTCAGGTGGTAACTACTTCTGCGGAGACACTGAGTTGTCTCA 962
QY 973 AGGGGATGACATGTTTTTGTGAACACCCATCATCTCTGAATTTACTTGAATAGTTAT 1032
Db 963 AGGGGATGACATGTTTTTGTGAACACCCATCATCTCTGAATTTACTTGAATAGT-TAT 1022
QY 1033 CCAAGACCAACCAAGATCTCTTCAATAGGTGATGATCCACAAATGCTGACCTCATC 1092
Db 1023 CCAAGACCAACCAAGATCTCTTCAATAGGTGATGATCCACAAATGCTGACCTCATC 1082
QY 1093 GGGATAAGAGCTGAAGTACTACATCCAGAGAGTTGTTTACTCTGCGCTCTACCAACCT 1152
Db 1083 GGGATAAGAGCTGAAGTACTACATCCAGAGAGTTGTTTACTCTGCGCTCTACCAACCT 1142
QY 1153 ATCGAGGGGAGGGGGTCTCTACTCTGCGGACACTGCGCTGAGGCTGACCATATCA 1212
Db 1143 ATCGAGGGGAGGGGGTCTCTACTCTGCGGACACTGCGCTGAGGCTGACCATATCA 1202
```

```
QY 1213 CTGACAGGTCCTCTACCCCATTTGATGACGCTTTTACTTGAAATGTCCTCAGAAAC 1272
Db 1203 CTGACAGGTCCTCTCTACCCCATTTGATGACGCTTTTACTTGAAATGTCCTCAGAAAC 1262
QY 1273 TCGGCTCGTTCAGAGAAACACAAAGGCTTCAGGACATTTTGATATCGAGGAGAAAAACA 1332
Db 1263 TCGGCTCGTTCAGAGAAACACAAAGGCTTCAGGACATTTTGATATCGAGGAGAAAAACA 1322
QY 1333 AAAATACATCTGCTCTCTATGTAGATCTGATGTTAGTACATAGTAGAAAACCTCAGAGA 1392
Db 1323 AAAATACATCTGCTCTCTATGTAGATCTGATGTTAGTACATAGTAGAAAACCTCAGAGA 1382
QY 1393 TGATTGATATTTGGTCTCAGTTCAGAGTGCCTCAITTTAAATGCTAAAATAGATACAAAC 1452
Db 1383 TGATTGATATTTGGTCTCAGTTCAGAGTGCCTCAITTTAAATGCTAAAATAGATACAAAC 1442
QY 1453 TCAATTTKGSATWGRAGGGGTTTGTGATWNGYCCCATTTTGGGCTCTCACAATAGAG 1512
Db 1443 TCAATTTGCAATAGAAAAGTGATTTTGAATAGTTCCTCATGTTGTTCTCATTAGAG 1502
QY 1513 TAATTTCTATTNAAC--ATGAAATTCCTTTTATGAGTGATACCCATTT-ANGGCTCT 1569
Db 1503 TAATTTCTATTAAACCATGAAATTTGCTTTATGAGTGATACCCATTTTGAGGGCTCT 1562
QY 1570 AANCCCTT--CATTTGNACTCAGCTGAAGGAAAGCGGAGAGAGTAAATTTTAT 1626
Db 1563 AANCCCTTCAATTTGGTACTCAGCTGAAG-AGGGAAGCGGAAGATGTTAATTTTTTT 1621
QY 1627 GGTGA 1631
Db 1622 ATGGA 1626
```

RESULT 5

```
AAF29255
ID AAF29255 standard; cdna; 1912 BP.
XX AAF29255;
XX AC AAF29255;
XX DT 17-APR-2001 (first entry)
XX DE Human beta 1,3-N-acetylglucosamine transferase G3 cDNA sequence.
XX KW Sugar chain synthesis agent; beta 1,3-N-acetylglucosamine transferase;
XX KW inflammation; cancer; metastasis; human; ss.
XX OS Homo sapiens.
XX PN WO200100848-A1.
XX PD 04-JAN-2001.
XX ZF 29-JUN-2000; 2000WO-JP04304.
XX PR 29-JUN-1999; 99JP-0183437.
XX PR 16-MAR-2000; 2000JP-0074757.
XX PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX PI Sasaki K, Shiraishi N, Natsume A, Yamada Y, Nakagawa S, Sekine S;
XX WPI; 2001-102895/11.
XX DR P-PSDB: AAB49748.
XX PT New polypeptide having beta1,3-N-acetylglucosamine transferase activity
XX PT for diagnosis of inflammation, cancer and cancer metastasis, -
XX PT development of remedies, and for producing glycoconjugates -
XX PS Example 2; Page 159-164; 195pp; Japanese.
XX CC This invention relates to a sugar chain synthesizing agent that contains
XX CC a polypeptide as the active ingredient, where the polypeptide has beta
XX CC 1,3-N-acetylglucosamine transferase activity. The polypeptides, DNA
```



```
FT sig_peptide 233..328
FT /*tag= b
FT mat_peptide 329..1423
FT /*tag= c
XX
XX W09844112-A1.
XX
XX 08-OCT-1998.
XX
XX 27-MAR-1998; 98WO-US06022.
XX
XX 28-MAR-1997; 97US-0042855.
XX
XX (HUMA-); HUMAN GENOME SCI INC.
XX
XX Ruben SM. Soppet DR;
XX
XX WPI; 1998-55711/47.
XX P-PSDB; AAW80212.
XX
XX New isolated cardiac and pancreatic protein - used to develop
XX products for the diagnosis and treatment of e.g. pancreatitis or
XX abnormal hypertrophy of the heart
XX
XX Claim 4: Fig 1; 113pp; English.
XX
XX This nucleotide sequence codes for novel cardiac and pancreatic
XX protein (CAPP, see AAW80212), a member of the muscle-derived growth
XX factor superfamily. It was discovered in a cDNA library derived
XX from activated T cells. The gene is abundant in adult heart and
XX pancreas, with low amounts in placenta, lung, liver, skeletal
XX muscle, kidney, spleen, thymus, testis, ovary, small
XX intestine, colon and peripheral blood leucocytes. CAPP cDNA is
XX deposited as ATCC 97729. A number of related sequences (see
XX AAV6374-405) are also provided. The invention relates to isolated
XX CAPP nucleic acid molecules and polypeptides, as well as vectors,
XX host cells and recombinant methods for producing the same. The
XX invention further relates to screening methods for identifying
XX agonists and antagonists of CAPP activity, as well as diagnostic
XX methods for detecting the presence of activated T-cells, and mature
XX heart, pancreas and placental tissues and cells. The CAPP
XX polypeptides can modulate the differentiation and proliferation of
XX cells and tissue, both in vivo and ex vivo. The products can be
XX used in the diagnosis and treatment of pancreatitis and conditions
XX that cause abnormal hypertrophy of the heart, such as hypertension,
XX myocardial infarction, valve disease and cardiomyopathy. The
XX products can also be used in detection and cell culturing.
XX
XX Sequence 2745 BP; 778 A; 526 C; 605 G; 836 T; 0 other;
XX
XX Query Match 87.24; Score 1488; DB 19; Length 2745;
XX Best Local Similarity 97.34; Pred. No. 0;
XX Matches 1575; Conservative 8; Mismatches 24; Indels 11; Gaps 7;
XX
XX 21 CGGCAGCAGCGGCAACAAAGTCCGCGAGGCTAGCAGACCCAGCGGAGCAGTCCCTGCCG 80
XX |||||||
XX 10 CGGCAGCAGCGGCAACAAAGTCCGCGA--CTAGCAGAGCCCAAGCCGAGCAGTCCCTGCCG 67
XX |||||||
XX 81 CCAGACACCGCGGCGCCCGCTCCGGGCGCGCGGATGGAGCGTGGAGTCCGCGGCTCG 140
XX |||||||
XX 68 CCGACACCGCGGCGCCCGCTCCGGGCGCGCGATGGAGCGTGGAGTCCGCGGCTCG 127
XX |||||||
XX 141 CC-GGGCTGAGCCCGCGGAGCGCGCGGACGTGGATGTGGCGCGGATCTCCGCGCCTTG 199
XX |||||||
XX 128 CCGGGGCTGAGCCGCGCGGAGC-GCCGGGACGTGGATGTGGCGCGCATCTCCCGCCTTG 186
XX |||||||
XX 200 CCCCCCGCCCGCGAGCTGAGCTGCTCCCGGACAGATATGAGAATGAGTGTGGACG 259
XX |||||||
XX 187 CCCCCCGCCCGCGAGCTGAGCTGCTCCCGGACAGATATGAGAATGAGTGTGGACG 246
XX |||||||
XX 260 TCGAAGAAATAAAGTTGTTGGGTATCCCTGATGATGGCAATGTCTTCAATTTATTAT 319
XX |||||||
XX 247 TCGAAGAAATAAAGTTGTTGGGTATCCCTGATGATGGCAATGTCTTCAATTTATTAT 306
XX |||||||
```

```
QY 320 GGAAGTCTCCAAAGCAGTAGCCCAAGAAAAAATGGAAGGGGAAGTAATAATACCAA 379
DB |||||||
DB 307 GGAAGTCTCCAAAGCAGTAGCCCAAGAAAAAATGGAAGGGGAAGTAATAATACCAA 366
QY 380 AGAGAAAGTTCTGGAAAGATATCTACCCCTCCCGAGGACATCTGGAAACCGAGAGCAAGAA 439
DB |||||||
DB 367 AGAGAAAGTTCTGGAAAGATATCTACCCCTCCCGAGGACATCTGGAAACCGAGAGCAAGAA 426
QY 440 CTTGAAACCGGAGTACAAACCCCATCTCTGAGCATGCTGACCAACAGAGAGGGGGAGGGG 499
DB |||||||
DB 427 CTTGAAACCGGAGTACAAACCCCATCTCTGAGCATGCTGACCAACAGAGAGGGGGAGGGG 486
QY 500 CAGGCTCTCCAAATATAAGCCATCTGAACCTACTGCGAACCTGACCTGAGGGTCACTCGGT 559
DB |||||||
DB 487 CAGGCTCTCCAAATATAAGCCATCTGAACCTACTGCGAACCTGACCTGAGGGTCACTCGGT 546
QY 560 GGTACAGGGTTTAAACAACCTTGGCGGACAGATTTAAAGACATTTCTGCTGTATTGAGATG 619
DB |||||||
DB 547 GGTACAGGGTTTAAACAACCTTGGCGGACAGATTTAAAGACATTTCTGCTGTATTGAGATG 606
QY 620 CGGCAATATTACCTGCTTATAGATCAGCGGATAGGTGTCGAAGAAACCTTTCTGTT 679
DB |||||||
DB 607 CGGCAATATTACCTGCTTATAGATCAGCGGATAGGTGTCGAAGAAACCTTTCTGTT 666
QY 680 GCTGCGGATTAAGTCCCTCACTCCACATTTTCCCAAGAGGCAAGCAATCCGGGAATCCTG 739
DB |||||||
DB 667 GCTGCGGATTAAGTCCCTCACTCCACATTTTCCCAAGAGGCAAGCAATCCGGGAATCCTG 726
QY 740 GGGCCAAAGAAACAGCAGGAGGAAACCAACGGTGGTGGAGTCTTCCTGCTGGGCCAGAC 799
DB |||||||
DB 727 GGGCCAAAGAAACAGCAGGAGGAAACCAACGGTGGTGGAGTCTTCCTGCTGGGCCAGAC 786
QY 800 ACCCCACAGGACAAACCAACCGGACCTTTTCAGATATGCTGAAATTTGAGAGTGAAGACA 859
DB |||||||
DB 787 ACCCCACAGGACAAACCAACCGGACCTTTTCAGATATGCTGAAATTTGAGAGTGAAGACA 846
QY 860 CCAAGACATTTCTATGTGGAACCTACAGAGACACTTTCTCAACTTGCTCTGTAAGAGAGT 919
DB |||||||
DB 847 CCAAGACATTTCTATGTGGAACCTACAGAGACACTTTCTCAACTTGCTCTGTAAGAGAGT 906
QY 920 GCTGTTCTCAGGTGGGTAAAGTACTTCTGCCAGACACTGAGTTTGTTCAGAGGGCGA 979
DB |||||||
DB 907 GCTGTTCTCAGGTGGGTAAAGTACTTCTGCCAGACACTGAGTTTGTTCAGAGGGCGA 966
QY 980 TGACCATGTTTGTGGAACCCCAACATCAGATCTGAATTTACTTGAATAGTTTATCAAGAC 1039
DB |||||||
DB 967 TGACCATGTTTGTGGAACCCCAACATCAGATCTGAATTTACTTGAATAGTTTATCAAGAC 1026
QY 1040 CAAAGCCAAAGATCTTCTATAGTGTGATGTGATCCACAAATGCTGGACCTCATTCGGGATAA 1099
DB |||||||
DB 1027 CAAAGCCAAAGATCTTCTATAGTGTGATGTGATCCACAAATGCTGGACCTCATTCGGGATAA 1086
QY 1100 GAAGCTGAAGTACTACATCCCAAGAGTTGTTTACTCTGSCCTCTACCCACCTTATGCGAGG 1159
DB |||||||
DB 1087 GAAGCTGAAGTACTACATCCCAAGAGTTGTTTACTCTGSCCTCTACCCACCTTATGCGAGG 1146
QY 1160 GGGAGGGGGTTCCTCTACTCCGGCCACTGSCCCTGAGGCTGTACCATATCACTGACCA 1219
DB |||||||
DB 1147 GGGAGGGGGTTCCTCTACTCCGGCCACTGSCCCTGAGGCTGTACCATATCACTGACCA 1206
QY 1220 GGTCCATCTCTACCCCATTTGAGCTTTTACTTGGAAATGCGCTTCAGAACTCGGGCT 1279
DB |||||||
DB 1207 GGTCCATCTCTACCCCATTTGAGCTTTTACTTGGAAATGCGCTTCAGAACTCGGGCT 1266
QY 1280 CTTCCAGAGAAACAAAGGCTTCAGGACATTTTATATCGAGGAGAAAAACAATAA 1339
DB |||||||
DB 1267 CTTCCAGAGAAACAAAGGCTTCAGGACATTTTATATCGAGGAGAAAAACAATAA 1326
QY 1340 CATCTGCTCTCTATGTAGATCTGATCTTAGTACATAGTAAAAACCTCAAGAGATGATGA 1399
DB |||||||
DB 1327 CATCTGCTCTCTATGTAGATCTGATCTTAGTACATAGTAAAAACCTCAAGAGATGATGA 1386
```

QY	1400	TATTTGGTCTCAGTTCGAGAGTGTCTCATTTAAAAAGCTAAATAGATACAAACTCAATTT	1459
Db	1387	TATTTGGTCTCAGTTCGAGAGTGTCTCATTTAAAAAGCTAAATAGATACAAACTCAATTT	1446
QY	1460	KGSATMGRAAGGGGTWTTTTGRATWGGYCCCATGTTGGGGCTCTCACATTAGAGTAATTTTC	1519
Db	1447	TGCATAGAAAGGTGATTTTGAATAGTTCCTCATGTTGTGTTCTCACATTAGAGTAATTTTC	1506
QY	1520	TATTTTNAAC--ATGAAATTCGCTTTTATGAGTGATACCCATTT-ANGGCCCTCTAAACGCTT	1576
Db	1507	TATATTAAACCATGAAAAATTCGCTTTTATGAGTGATACCCATTTGAGGGCCTCTAAACCCCT	1566
QY	1577	---CATTTGNACTCACGCTGAAGAGGCGAAGCGGAGAGGTAATTTTATGCGTGA	1631
Db	1567	TCATTTGGTACTCACGCTGAG-AGGAAAGCGGAAGATGTAATTTTTTTTATGGA	1623
RESULT 7			
AAH14972			
ID	AAH14972	standard; cDNA; 2442 BP.	
XX	AAH14972;		
XX	26-JUN-2001	(first entry)	
XX	Human	cDNA sequence SEQ ID NO:12897.	
XX	Human;	primer; detection; diagnosis; antisense therapy; gene therapy; ss.	
XX	Homo sapiens.		
XX	EP1074617-A2.		
XX	07-FEB-2001.		
XX	28-JUL-2000;	2000EP-0116126.	
XX	29-JUL-1999;	99JP-0248036.	
XX	27-AUG-1999;	99JP-0300253.	
XX	11-JAN-2000;	2000JP-0118776.	
XX	02-MAY-2000;	2000JP-0183767.	
XX	09-JUN-2000;	2000JP-0241899.	
XX	(HELI-)	HELIX RES INST.	
XX	Ota T. Isogai T. Nishikawa T. Hayashi K. Saito K. Yamamoto J.		
XX	Ishii S. Sugiyama T. Wakamatsu A. Nagai K. Otsuki T.		
XX	WPI: 2001-318749/34.		
DR			

CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AB92446 to
CC AB95993 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination
CC of the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by


```

Dd 508 AGCAACGAGGAAACCAACGCTGGTGGAGTCTCTGCTGGCCAGACACCCCCAGAG 567
Qy 810 GACAACACCCCGACCTTTGAGATATCTGTAATTTGAGAGTGAAGACCAAGACAT 859
Dd 568 GACAACACCCCGACCTTTGAGATATCTGTAATTTGAGAGTGAAGACCAAGACAT 627
Qy 870 CTTATGTGGAACTACAGACACTTTCTTCAACTTGTCTGGAAGGAGTCTGTTTC 929
Dd 528 CTTATGTGGAACTACAGACACTTTCTTCAACTTGTCTGGAAGGAGTCTGTTTC 687
Qy 930 AGTGGTAACTACTTCTGCCACACACTGAGTTTCTTCAAGGGGATGACGATGT 989
Dd 688 AGTGGTAACTACTTCTGCCACACACTGAGTTTCTTCAAGGGGATGACGATGT 747
Qy 990 TTTGTGAACCCCATCATCTCTGAATTTACTTGAATAGTTTATCCAAAGCCAAAGCCAAA 1049
Dd 748 TTTGTGAACCCCATCATCTCTGAATTTACTTGAATAGTTTATCCAAAGCCAAAGCCAAA 807
Qy 1050 GATCTCTTATAGTGTGATCCCAATGCTGGACCTCATCGGGATAGAAGCTGAAG 1109
Dd 808 GATCTCTTATAGTGTGATCCCAATGCTGGACCTCATCGGGATAGAAGCTGAAG 867
Qy 1110 TACTACATCCAGAGTTGTTTACTCTGGCTCTACCCACCTATGCGAGGGGAGGGGG 1169
Dd 868 TACTACATCCAGAGTTGTTTACTCTGGCTCTACCCACCTATGCGAGGGGAGGGGG 927
Qy 1170 TTCTCTACTCGGCCACCTGGCCCTGAGCTGTACCATATCACTGACGAGTCCATCTC 1229
Dd 928 TTCTCTACTCGGCCACCTGGCCCTGAGCTGTACCATATCACTGACGAGTCCATCTC 987
Qy 1230 TACCCCATGATGACGTTTATCTGGAATGTGCTTCAGAAACTCGGCCCTGTTCCAGAG 1289
Dd 988 TACCCCATGATGACGTTTATCTGGAATGTGCTTCAGAAACTCGGCCCTGTTCCAGAG 1047
Qy 1290 AAACACAAGCCTTCAGACATTTGATATCGAGAGAAAACAAAATATACATCTGCTCC 1349
Dd 1048 AAACACAAGCCTTCAGACATTTGATATCGAGAGAAAACAAAATATACATCTGCTCC 1107
Qy 1350 TATGTAGATCTGATTTAGTACATAGTAGAAACCTCAAGAGATGATTGATTTGGTCT 1409
Dd 1108 TATGTAGATCTGATTTAGTACATAGTAGAAACCTCAAGAGATGATTGATTTGGTCT 1167
Qy 1410 CAGTTGACAGTCTCATTTAAATGCTAAATAGATACAAACCTCAATTTKGSATWGRAA 1469
Dd 1168 CAGTTGACAGTCTCATTTAAATGCTAAATAGATACAAACCTCAATTTGTCATAGAAA 1227
Qy 1470 GGGGTWTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1529
Dd 1228 GGGGTWTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1287
Qy 1530 --ATGAAATTCCTTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1583
Dd 1288 CATGAAATTCCTTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1347
Qy 1584 ACTCAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1531
Dd 1348 ACTCAGTGAAG-AGGGAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1394

```

RESULT 9

AAZ88487

ID AA288487 standard; cDNA; 1194 BP.

XX

AC

XX

DT 12-MAY-2000 (first entry)

XX

DE Human brainiac protein encoding cDNA sequence.

XX

KW Brainiac; egghad; cell adhesion; cytoskeletal; dermatological;

KW neuroactive; cell viability; cell growth regulation; psoriasis; cancer;

XX cell fate specification; skin lesion; nervous system defect; ss.

QS Homo sapiens.

XX CA2225126-A1.

PN XX

PD 17-JUN-1999.

XX

PF 17-DEC-1997; 97CA-2225126.

XX

PR 17-DEC-1997; 97CA-2225126.

XX

PA {HSCR-} HSC RES & DEV LP.

XX

PI Egan SE;

XX

DR WPI: 2000-161481/15.

XX

DR P-PSDB; AAY79954.

XX

Mammalian EGGHEAD and BRAINAC proteins which mediate cell to cell adhesion and may be used to treat cancer, psoriasis and other skin lesions and nervous system defects or diseases -

XX

PS Example 2; Page 27; 30pp; English.

XX

The present invention describes mammalian EGGHEAD and BRAINAC proteins, which mediate cell-to-cell adhesion. The mammalian proteins are similar to the EGGHEAD and BRAINAC genes of *Drosophila melanogaster*. BRAINAC and EGGHEAD genes regulate adhesion between epithelial cells and require the presence of the Notch protein for this activity. Mutational disruption of BRAINAC, EGGHEAD or Notch results in the loss of follicle epithelial cell adhesion, thereby reducing the efficiency of signalling through other epithelial cell receptors such as the epidermal growth factor receptor. The BRAINAC and EGGHEAD proteins have cytoskeletal, dermatological and neuroactive properties. As the BRAINAC and EGGHEAD proteins are involved in adhesion between epithelial cells, and as this adhesion is required for cell viability, cell growth regulation and cell fate specification, it is envisioned that wild type or mutant forms of mammalian BRAINAC and/or EGGHEAD can be used to alter epithelial cell adhesion. This should be useful in treating many diseases which present problems of cell viability, cell growth regulation and cell fate specification. For example, these proteins, or active fragments or analogues of these proteins and these genes can be used to treat diseases such as cancer, psoriasis and other skin lesions, and nervous system defects or diseases. The present sequence encodes the human BRAINAC protein.

XX

SQ Sequence 1194 BP; 312 A; 242 C; 243 G; 259 T; 138 other;

Query Match 60.8%; Score 1039.4; DB 21; Length 1194;

Best Local Similarity 89.5%; Pred. No. 8.3e-274;

Matches 1045; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

Qy 273 TTGTTGGTATCTGTATGATGCAATATCTTCATTTATTTATTTATGAGTCTCCAA 332

Dd 28 TTCTGGGATCTGTATGATGCAATATCTTCATTTATTTATTTATGAGTCTCCAA 87

Qy 333 AGCAGTAGCCAGCAAGAAAAATGAAAGGGGAGTAATATACCCAAAGAGTCTCG 392

Dd 88 AACAGTAGCCAGCAAGAAAAATGAAAGGGGAGTAATATACCCAAAGAGTCTCG 147

Qy 393 AGATATATACCCCTCCGAGGCAATGTAACCCAGAGAGAGAGTGAACCGGAG 452

Dd 148 AGATATATACCCCTCCGAGGCAATGTAACCCAGAGAGAGAGTGAACCGGAG 207

Qy 453 TACAACCCCATCTGTAGCATGTGACCAACAGAGGGGAGGCGGAGGCTCTCCAT 512

Dd 208 TACAACCCCATCTGTAGCATGTGACCAACAGAGGGGAGGCGGAGGCTCTCCAT 267

Qy 513 ATAAGCCATCTGAACCTACTGCGAACCTGACCTGAGGTCACGTCGCTGCTTACGGT 572

Dd 268 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 327

Qy 573 AACAACTTCCCGACAGATTTAAAGACTTTCTGCTCTATTGAGATGCCCAATATTCA 632

```
Db 328 AACAACTGCGGACAGATTAAAGACTTTCTGCTGATTTAGATGCCGAATATTCA 387
Qy 633 CTGCTTAGATACAGCGGATAAGTGTGCAAGAAACCTTCTTGCTGGCGATTAG 692
Db 388 CTGCTTAGATACAGCGGATAAGTGTGCAAGAAACCTTCTTGCTGGCGATTAG 447
Qy 693 TCCCTCACTCCACATTTTGCAGAAAGCAAGCAATCCCTGGGCCAAGAAAGC 752
Db 448 TCCCTCACTCCACATTTTGCAGAAAGCAAGCAATCCCTGGGCCAAGAAAGC 507
Qy 753 AACCCAGGAAACCAACCGTGGCGAGTCTTCTGCTGGCCAGACACCCCGAGGAC 812
Db 508 AACCCAGGAAACCAACCGTGGCGAGTCTTCTGCTGGCCAGACACCCCGAGGAC 567
Qy 813 AACCCACCCGACCTTTTCAGATATGCTGAAATTTGAGAGTGAGAGCAACCAAGACATCTT 872
Db 568 AACCCACCCGACCTTTTCAGATATGCTGAAATTTGAGAGTGAGAGCAACCAAGACATCTT 627
Qy 873 ATGTGGAACTACAGACACTTTCTCAACTTGTCTGAGGAAGTGTCTTCTCAGG 932
Db 628 ATGTGGAACTACAGACACTTTCTCAACTTGTCTGAGGAAGTGTCTTCTCAGG 687
Qy 933 TGGTAAAGTACTTCTCCAGACACTGAGTGTGTTTCAAGGGCGATGACGATGTTT 992
Db 688 TGGTAAAGTACTTCTCCAGACACTGAGTGTGTTTCAAGGGCGATGACGATGTTT 747
Qy 993 GTGAACACCCATCACATCTGAAATTTACTTGAATAGTTTATCCAGACCAAGCAAGAI 1052
Db 748 GTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 807
Qy 1053 CTCTTCATAGTGATGTATCCACAATCTGGACCTCATCGGATAGAAAGCTGAAGTAC 1112
Db 808 CTCTTCATAGTGATGTATCCACAATCTGGACCTCATCGGATAGAAAGCTGAAGTAC 857
Qy 1113 TACATCCAGAAGTGTGTTTACTCTGGCTCTACCCACCTATCAGGGGAGGGGCTC 1172
Db 868 TACATCCAGAAGTGTGTTTACTCTGGCTCTACCCACCTATCAGGGGAGGGGCTC 927
Qy 1173 CTCTACTCGGCGACCTGGCCCTGAGGCTGTACCATATCACTGACCAAGTCTCTAC 1232
Db 928 CTCTACTCGGCGACCTGGCCCTGAGGCTGTACCATATCACTGACCAAGTCTCTAC 987
Qy 1233 CCCATGATGACGTTTATCTGGAATGTGCTTCAGAACTCGGCTCTGTTCCAGAA 1292
Db 988 CCCATGATGACGTTTATCTGGAATGTGCTTCAGAACTCGGCTCTGTTCCAGAA 1047
Qy 1293 CACAAAGGCTTCAGGACATTTGATATCAGGAGAAACCAAAATACATCTGCTCTAT 1352
Db 1048 CACAAAGGCTTCAGGACATTTGATATCAGGAGAAACCAAAATACATCTGCTCTAT 1107
Qy 1353 GTAGATCTGATGTTAGTACATAGTAGAAACCTTCAAGAGATGATGATTTGGTCTAG 1412
Db 1108 GTAGATCTGATGTTAGTACATAGTAGAAACCTTCAAGAGATGATGATTTGGTCTAG 1167
Qy 1413 TTGCAGAGTGTCTCAITTTAAATGCTAA 1439
Db 1168 TTGCAGAGTGTCTCAITTTAAATGCTAA 1194
```

RESULT 10

AAZ88486

ID AAZ88486 standard; cDNA; 2676 BP.

XX

AAZ88486;

XX

12-MAY-2000 (first entry)

XX

Murine brainiac encoding cDNA sequence.

XX

Brainiac; egghed; cell adhesion; cytoskeletal; dermatological;

XX neuroactive; cell viability; cell growth regulation; psoriasis; cancer;

XX cell fate specification; skin lesion; nervous system defect; ss.

XX

Mus sp.

XX CA2225126-A1.

PN

XX 17-JUN-1999.

PD

XX 17-DEC-1997; 97CA-2225126.

PF

XX 17-DEC-1997; 97CA-2225126.

PR

XX (HSCR-) HSC RES & DEV LP.

PA

XX Egan SE;

XX

XX WP: 2000-161481/15.

DR

XX P-PSDB; AAY79953.

DR

XX Mammalian EGGHEAD and BRAINIAC proteins which mediate cell to cell adhesion and may be used to treat cancer, psoriasis and other skin lesions and nervous system defects or diseases.

PT

XX Example 1; Page 24-25; 30pp; English.

PS

The present invention describes mammalian EGGHEAD and BRAINIAC proteins, which mediate cell-to-cell adhesion. The mammalian proteins are similar to the EGGHEAD and BRAINIAC genes of Drosophila melanogaster. BRAINIAC and EGGHEAD genes regulate adhesion between epithelial cells and require the presence of the Notch protein for this activity. Mutational disruption of BRAINIAC, EGGHEAD or Notch results in the loss of follicle epithelial cell adhesion, thereby reducing the efficiency of signalling through other epithelial cell receptors such as the epidermal growth factor receptor. The BRAINIAC and EGGHEAD proteins have cytoskeletal, dermatological and neuroactive properties. As the BRAINIAC and EGGHEAD proteins are involved in adhesion between epithelial cells, and as this adhesion is required for cell viability, cell growth regulation and cell fate specification, it is envisioned that wild type or mutant forms of mammalian BRAINIAC and/or EGGHEAD can be used to alter epithelial cell adhesion. This should be useful in treating many diseases which present problems of cell viability, cell growth regulation and cell fate specification. For example, these proteins, or active fragments or analogues of these proteins and these genes can be used to treat diseases such as cancer, psoriasis and other skin lesions, and nervous system defects or diseases. The present sequence encodes the murine BRAINIAC protein.

XX

SQ Sequence 2676 BP; 732 A; 562 C; 637 G; 744 T; 1 other;

Query Match 58.7%; Score 1002.2; DB 21; Length 2676;

Best Local Similarity 83.0%; Pred. No. 11e-263;

Matches 1195; Conservative 7; Mismatches 225; Indels 12; Gaps 5;

Qy 79 CGCCGACACCCGCGCGCCGCTCCGGGGCCCGCGATGGAGCTGGCGCGGT 138

Db 53 CCAGACCCCGCGCGCGCGCTCCCGCGATGGAGCTGGCGCGGT 110

Qy 139 CGCGGGGTGACCGCGCGGAGCGCGCGGAGCGTGGATGTGGCGCGATCTCCGCGCT 198

Db 111 CGCGGGGTGACCGCGGAGCGCGCGGAGCGTGGATGTGGCGCGATCTCCGCGCT 164

Qy 199 GCGCGCGCGCGCGCGCGCGCTGGAGCTGGCTCCCGGACAAATGAGAAATGAGTGGAC 258

Db 165 --CCCGCGCGCGCGCGCGCGCTGGAGTGTCCCTAGACAAAGTATGAGAGATGAGTGGGCG 222

Qy 259 GTCGAGAGATAAAGTGTGGTGTATCTGATGATGCGCAATGCTCTTCATTTATTATTA 318

Db 223 GTCGAGAGACTCAAGTGTGGCATCTGATGATGCGCAATGCTCTTCATTTATTATTA 282

Qy 319 TGGAGTCTCCAAAAGCAGTAGCCAGAAAGAAATGGAAGAGGAGAGTAATATACCA 378

Db 283 TGGAGTCTCCAAAAGCAGTAGCCAGAAAGAAATGGAAGAGGAGAGTAATATACCA 342

Qy 379 AGAGAGTCTTGGAGATATCTACCCCTCCGAGGCGATCTGGAGAGGAGAGAGAG 438

XX

Db 343 AAGAGAAGTTCTGGAAGCCACCCAGCAGCCTCCCGGGCATACIGGAACAGGGAACAGGAGA 402
Qy 439 AGCTGAACCGGCAGTACAAACCCATCTGAGCATGCTGACCAACAGCGGGGAGCGG 498
Db 403 AGCTGAACAGGTGTGTAATCCCATCTTGACAGAGGTGGCCATCAGACAGGGGAGGTAG 462
Qy 499 GAGGCTCTCAATATAGGCATCTGAACCTGCTGACCTGACCTGAGGTGACGTGCG 558
Db 463 CCACATCTCCAAACACAGTCACTGAGCTATGTTGAACACAGATCGACGTCATGACAG 522
Qy 559 TGGTTACGGTTTTAAACATGTCGGACAGATTTAAAGACTTTCTGCTGTATTGAGAT 618
Db 523 CTGTGACAGATTTTAAATAATCTGCGGACAGATTTAAAGACTTTCTGCTGTATTGAGAT 582
Qy 619 GCCCAATTATTCACTGCTATAGATCAGCGGATAGTGTGGAAGAAACCTTTCTTGT 678
Db 563 GCCGGAATTACTGCTGCTTATAGATCAACCGAAGAAATGTGAAAGAACGCCCTTCTTAC 642
Qy 679 TGCTGGGGATTAAAGTCCCTCACTCCACATTTTGCAGAGCAAGCAATCGGGAATCCT 738
Db 643 TATTGGCGATAAGTCCCTCATTCACATTTTGCAGAGCAAGCAATTCGGGAGTCTT 702
Qy 739 GGGGCCAAGAAAGCAAGCGAGGAAACCAAGCGTGGTGGAGTCTTCTGCTGGGCCAGA 798
Db 703 GGGGCCGAGAAACCAAGCTAGGGAACAGACAGTGTGAGGGTCTTCTGTTGGGCAAGA 762
Qy 799 CACCCCGAGAGGACCAACCCACCTTCAGATATGCTGAATTTGAGAGTGAGAGC 858
Db 763 CACCCCGAGAGGACCAACCCACCTTCAGATATGCTGAATTTGAGAGTGAGAGC 822
Qy 859 ACCAAGACATCTTATGTGAACATACAGACACATTTCTCAACTTGTCTCTGAAGGAAG 918
Db 823 ACCAGGACATCTCATGTGGAACATATAGACACATTTCTCAACTGTCCCTGAAGGAAG 882
Qy 919 TGCTGTTTCTCAGTGGTAAAGTACTTCTGCCAGACACATGAGTTTGTTCGAAGGCG 978
Db 883 TGCTGTTTCTTAGTGGGTGAGCACTTCTCTGCCAGCGCAGAGTTTGTCTCAAGGCG 942
Qy 979 ATGACGATGTTTTGTGAACACCATCACATCCTGAATTTACTGAATGTTATCCAAGA 1038
Db 943 ATGATGAGCTGTTGTGACACCCATCACATCCTTAATTTACTTGAATGCTTATCCAAGA 1002
Qy 1039 CCAAGCCAAAGATCTCTTCATAGGTGATGTGATCCCAATGCTGTGACCTCATCGGATA 1098
Db 1003 GCAAGCCAAAGACTTGTTCATAGTGTGACGTGATCCCAATGCTGTGACCTCATCGGATA 1062
Qy 1099 AGAAGCTGAAGTACTATCCCAAGATGTTTACTCTGGCTCTACCCACCTATGCG 1158
Db 1063 AGAACTGAAGTACTATCCCAAGATGTTTACTCTGGCTCTACCCACCTATGCG 1122
Qy 1159 GGGGAGGGGGTTCCTCTACTCCGGCCACCTGGCCCTGAGGCTGTACATATCACTGACC 1218
Db 1123 GGGGTGTGGATCTCTACTCCGGCCCTTGTCTGAGGCTGTACGTCGACTAGCC 1182
Qy 1219 AGTCCATCTCTACCCATGATGACGTTTATCTGGAATGTGCTTCAGAAACTCGGCC 1278
Db 1183 GGGTCCATCTCTACCCATGATGATGTTTATACGGGAATGTGCTTCAGAAACTCGGCC 1242
Qy 1279 TCGTTCAGAGAAACACAAAGGCTTCAGACATTTGATTCGAGGACAAACAAATA 1338
Db 1243 TTGTCAGAGAGACCAAAAGGCTTCAGGACATTTGATTTGAACAGAAATAAAGAAA 1302
Qy 1339 ACATCTGCTCTATGTAGATCTGATGTTAGTACATAGTAAAAACCTCAAGAGATGATTG 1398
Db 1303 ATATTGTTCTATATAGACCTAATGTTAGTACATAGCAAAACCTCAAGAGATGATTG 1362
Qy 1399 ATATTGTTCTAGTTCAGAGTCTCATTTAAATGCTAAAAATAGATACAAACTCAATT 1458
Db 1363 ATATCTGGTCTCACTTCAAAAGTCTTAATTTAAATGCTGAATAGACATGAGCTGATT 1422
Qy 1459 TKSATWGRAGGGGTTTWTGRTWGGGCCATGTTGGGTCTCACCATTAGAGTAATT 1517
Db 1423 T--CACAGAAAGGCCTAGCTGACTAGTTCCTAGTGTGCTCTCAACAATAGGTGAGTT 1479

RESULT 11
AAZ87185
ID AAZ87185 standard; cDNA; 2676 BP.
XX
AC AAZ87185.
XX
DI 08-MAY-2000 (first entry)
XX
DE Murine Brainiac cDNA.
XX
KW Brainiac; murine; mammalian; expressed sequence tag; EST;
KW Drosophila melanogaster; Egghead; Notch; epithelial cell adhesion;
KW Fringe family; cell viability; growth regulation; cell fate; cancer;
KW psoriasis; skin lesion; nervous system disorder; developmental syndrome;
KW transgenic animal; drug screening; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
CDS 210..1403
FT FT /*tag= a
FT FT /product= "Murine Brainiac protein"
XX
PN CA2255109-A1.
XX
PD 17-JUN-1999.
XX
PF 17-DEC-1998; 98CA-2255109.
XX
PR 17-DEC-1997; 97CA-2255126.
XX
PA (HSCR-) HSC RES & DEV LP.
XX
XX Egan SE:
XX
DR WPI: 2000-148082/14.
DR P-PSDB; AAY69697.
XX
PT New nucleic acids encoding a murine and human Brainiac protein, useful
PT for detecting somatic or germline DNA lesions which are responsible for
PT developmental syndromes or diseases including cancer .
XX
PS Claim 5; Page 26-27; 40pp; English.
XX
CC This sequence represents cDNA encoding murine Brainiac protein. The cDNA
CC was isolated from a mouse mammary gland cDNA library via the use of a
CC probe generated via PCR from a variety of murine tissues. The PCR
CC primers used for probe generation (AAZ87187-287188) were based on EST
CC (expressed sequence tag) sequences with significant similarity to
CC Drosophila melanogaster Brainiac. Drosophila Brainiac and Egghead
CC proteins regulate adhesion between epithelial cells, this activity being
CC dependent on the presence of Notch. Drosophila Brainiac is a secreted
CC protein, and has sequence similarities with the Drosophila Fringe
CC for cell viability, cell growth regulation and cell fate specification.
CC Wild-type or mutant forms of mammalian Brainiac proteins may therefore be
CC used to alter epithelial cell adhesion in a mammal. Mammalian Brainiac
CC proteins, active fragments analogues, and nucleic acids may be used to
CC treat diseases such as cancer, psoriasis and other skin lesions, and
CC nervous system disorders. Mammalian Brainiac nucleic acids may also be
CC used to detect somatic or germline DNA lesions which are responsible for
CC developmental syndromes or diseases including cancer. The mammalian
CC Brainiac proteins and fragments or its analogues are useful as antigens
CC in immunoassays including enzyme-linked immunosorbent assays (ELISA),
CC radioimmunoassays (RIA) and other non-enzyme linked antibody binding
CC assays. Non-human transgenic animals comprising nucleotide sequences
CC encoding human Brainiac protein(AAY69698) can be used as animal models
CC for the study of mammalian Brainiac gene function, for the screening of
CC candidate compounds and for the evaluation of potential therapeutic
CC interventions.
XX

CC The invention relates to a transgenic mouse comprising a disruption in a
 CC beta3gnf (beta-1,3-N-acetylglucosaminyltransferase) gene, where there is
 CC no native expression of beta3gnf gene. The transgenic mice may be used
 CC as in vivo model to study various disease states or conditions in which
 CC beta3gnf may be implicated or involved, such as abnormal cell growth,
 CC cancer and metastasis, and to evaluate various treatments or to identify
 CC agents for treating disease states or conditions, such as anxiety or
 CC depression. Animal based disease systems may be used to identify
 CC compounds capable of ameliorating disease symptoms, as test substrates
 CC for the identification of drugs, pharmaceuticals, therapies and
 CC interventions that may be effective in treating a disease or other
 CC phenotypic characteristic of the animal. The present sequence represents
 CC a mouse beta3gnf polynucleotide sequence.

xx Sequence 1260 BP; 370 A; 282 C; 307 G; 301 T; 0 other;

Query Match 54.6%; Score 931.4; DB 25; Length 1260;
 Best Local Similarity 84.1%; Pred. No. 1.8e-244;
 Matches 1049; Conservative 1; Mismatches 197; Indels 0; Gaps 0;

QY 217 TGGAGCTGCTCCGGCAAGCATATGGAATGACTGTTGGACCTCGAAGATAAGTTGT 276
 Db 1 TGGAGGTGCTCCCTAGACAGGTATGAGATGAGTGTGGGCGCTCGAAGTCAAGTTGC 60

QY 277 TGGGTATCTGATGCGCAATGTCTTCATTTATTTATTTATGGAAGTCTCCAAAGCA 336
 Db 61 TGGCATCTCTGATGCGCAATGTCTTCATTTATTTATTTGTTGGAAGTCTCCAAAGCA 120

QY 337 GTAGCCCAAGAAAAATGGAAGGGGAAGTATATATACCAAGAGAGTCTCTGAAGA 396
 Db 121 GTAGCCCAAGAAAAATGGAAGGGGAGGAGTATATATCCCGAAGAGAGTCTCTGAAGC 180

QY 397 TATCTACCTCCCTCCGGGCACTACTGGAACCGAGAGCAAGAGAGTCAACCGCAGTACA 456
 Db 181 CACCCAGCACTCCCGGGCATACTGGAACGGGACAGAGAGAGTCAACAGGTGGTACA 240

QY 457 ACCCCATCTGAGATGCTGACCAACGAGCGGGGAGGCGGAGGCTCTCCAAATAA 516
 Db 241 ATCCCATCTTGAACAGGCTGGCCAAATCAGACAGGGGAGCTAGCCACATCTCCAAACAA 300

QY 517 GCCATCTGAATCTGCAACCTGACCTGAGGTGCTGCTGGTGTACGGGTTTTAACA 576
 Db 301 GTCACCTGAGCTATTTGAACAGCAGCTGACGGTCTGACAGCTGTGACAGATTTAATA 360

QY 577 ACTTCCGGGACAGATTTAAAGACTTTCTGCTGTATTTGAGATGCCGAATTTATCACTG 636
 Db 361 ATCTGCCGACAGATTTAAAGACTTTCTCTGTATTTGAGATGCCGAATTTACTGCTGC 420

QY 637 TTATAGATCAGCCGGATAAGTGTGCAAGAAACCTTTCTTGTGCTGGCGATTAAAGTCC 696
 Db 421 TTATAGATCAACCGAAGAAATGTGCAAGAAAGCCCTTCTTACTATTGGCGATAAGTCC 480

QY 697 TCATCCACATTTGCCGAGAGCAAGCAATCCGGGAATCTCGGGCCAGAGAGCAAGC 756
 Db 481 TCATTCACATTTTGCCGAGAGCAAGCAATTCGGGAGTCTTGGGGCCGAGAAACCAAGC 540

QY 757 CAGGGAACCAACCGGTGTGGAGCTTCTCTGCTGGGCGAGACACCCCGAGAGCAAC 816
 Db 541 TAGGGAACCAACAGTGTGAGGCTCTCTCTGTTGGGCGAGACACCCCGAGAGCAAC 600

QY 817 ACCCGGACCTTTCAGATATGCTGAAATTTGAGAGTGAAGACCAAGACATTTATGT 876
 Db 601 ACCCTGACCTTTCGAGATGCTTAAAGTTTGAAGTGAACAGCACCAGGATCTCTATGT 660

QY 877 GGAATCAGAGACACTTTCTCAACTGTCTCTGAAGAGAGTCTGTTCTCAGGTGG 936
 Db 661 GGAATCAGAGACACTTTCTCAACTGTCTCTGAAGAGAGTCTGTTCTCAGGTGG 720

QY 937 TAAGTACTTCTGCCAGACACTGAGTTTGTTCAGGGGAGTGAAGTCTTTCTGTA 996
 Db 721 TGAGACTTCTGTCACAGCAGAGTTTGTTCAGGGGAGTGAAGTCTTTCTGTA 780

QY 997 ACACCCATCATCTCTGAATTTACTTGAATAGTTTATCCAGACCAAGCAAGATCTCT 1056

Db 781 ACACCCATCATCTCTTAATTTACTTTGATAGCTTAATCCAGAGCAAGCAAGACTTGT 840
 QY 1057 TCATAGGTGATGTATGCCAATATGCTGGACCTCATCGGGATAAGAGTGAAGTACTACA 1116
 Db 841 TCATAGGTGACGTGATGCCAATATGCTGGGCTCTCACTCGGATAAGAACTGAAGTACTACA 900
 QY 1117 TCCAGAGTGTCTTACTCTGGCCTCTACCCACCTATGACAGGGGAGGGGTTCTCTCT 1176
 Db 901 TCCAGAGTGTCTTACTACACCGCGCTATCCCACTATGCGGGGTTGGTGAATCTCTGT 960
 QY 1177 ACTCGGGCCACCTGCGCCTGAGGCTGTACCATATCACTGACAGGTCCATCTCTACCCCA 1236
 Db 961 ACTCGGGCCCTTGGCTTGAAGCTGTACAGTGCAGCTAGCGGGTCCATCTCTACCCCA 1020
 QY 1237 TTGATGAGCTTTATCTGGAATGCTGCTTCAGAACTCGGCTCGTTCCAGAGAAACACA 1296
 Db 1021 TTGATGATGTTTATACGGGAATGTGCTTCAGAACTGGGCTTGTTCAGAGAAACACA 1080
 QY 1297 AAGGCTTCAGGACATTTGATATCGAGGAGAAACAAACAAATTAACATCTGCTATGTAG 1356
 Db 1081 AAGGCTTCAGGACATTTGATATGAGAGAAACAAATTAACAAATATTGTTCTCTATAG 1140
 QY 1357 ATCTGATGTTAGTACATAGTAGAAAACTCAAGAGATGATGATATTGGTCTCAGTTGC 1416
 Db 1141 ACCTAATGTTAGTACATAGCAGAAAACTCAAGAGATGATGATATCTGGTCTCAGTTGC 1200
 QY 1417 AGAGTGCCTCAATTAATGCTAAATAGATATACAACTCAATTTKGA 1463
 Db 1201 AAGTCTCAATTAATGCTAAATAGATATGATGAGCTGATGCTCAATTTTACA 1247

RESULT 13
 AA12714
 ID AA12714 standard; cDNA; 1196 BP.
 AC AA12714;
 DT 25-JUL-2000 (first entry)
 XX cDNA encoding murine LIG46 polypeptide.
 XX Leptin; LIG46; body weight; leptin inducible gene; obesity; cachexia;
 KW LIG56; Tgtp; LRG-47; RC10-II; Stral3; ss.
 CS Mus sp.
 XX Key Location/Qualifiers
 FT CDS 3..1196
 FT /tag= a
 FT /product= "LIG46"
 XX W0200015826-A2.
 XX 23-MAR-2000.
 XX 10-SEP-1999; 99WO-US20722.
 XX 10-SEP-1998; 98US-0150857.
 PR 29-OCT-1998; 98US-0106378.
 PR 19-NOV-1998; 98US-0195896.
 PR 15-APR-1999; 99US-0292228.
 XX (MILJ-) MILLENNIUM PHARM INC.
 XX White B, Zhou J, Tartaglia LA;
 DR WPI; 2000-271461/23.
 DR P-PSDB; AAY84641.
 XX Method for determining compounds which modulate body weight and can be
 PT used to treat e.g. obesity comprises measuring the activity of leptin
 PT inducible genes

XX

PS Claim 16; Fig 1A-B; 123pp; English.

XX The present sequence encodes a leptin induced LIG46 polypeptide. The
XX specification describes a method for determining whether a compound can
XX be used to modulate body weight by measuring the activity of leptin
XX inducible genes, such as LIG46. The method can be used to specifically
XX detect LIG46 and LIG56 nucleic acid molecules relative to other nucleic
XX acid molecules encoding galactosyltransferases or GTP-binding proteins.
XX They can also be used in diagnostic assays to identify the presence or
XX absence of a genetic lesion or mutation characterized by aberrant
XX modification or mis-regulation of the genes or aberrant post-
XX translational modification of the proteins. LIG46 and LIG56 proteins
XX and nucleic acid molecules can be used to treat obesity or cachexia.
XX LIG46 and LIG56 antagonists are used to treat obesity and LIG46 and
XX LIG56 agonists are used to treat low body weight. The leptin inducible
XX genes i.e. LIG46, LIG56, Tgtp, LRG-47, RC10-II and Stral3 can be used
XX to determine whether a compound modulates body weight and can then be
XX used to treat obesity or cachexia or low body weight.

SQ Sequence 1196 BP; 350 A; 269 C; 290 G; 287 T; 0 other;

Query Match 53.1%; Score 906.4; DB 21; Length 1196;
Best Local Similarity 84.9%; Pred. No. 1.2e-237;

Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 244 AAATGACGTGTGGAGCTCGAAGATAAAGTTGTGGGTATCCCTGATGATGCGAATGTCT 303
DB 1 AGATGAGTGTGGGGCGTCGAAGAGTCAAGTTGTGGGCATCCCTGATGCGAATGTCT 60
QY 304 TCATTATTTTATTATTAAGCTCTCCAAAGACAGTCCCAAGAAAATAATGGAAAAGGG 363
DB 61 TCATTATTTGATGTGGAGCTCTCCAAARACAGTCCCAAGAAAATAATGGAAAAGGG 126
QY 364 AAGTAATATACCAAGACAGAGTTCTGGAGATATATCTACCCCTCCCGAGGCACTGGA 423
DB 121 GAGTAATATCCGAAGAGAGAGTTCTGGAAAGCCACCCAGCAGCTCCCGGGGCACTGGA 180
QY 424 ACCGAGACAGAGAGAGCTGAACGGGAGTACACACCCCATCTCGAGCATGCTGACCAACC 483
DB 181 ACAGGACAGAGAGAGCTGACAGAGTGTGACATATCCATCTTGAACAGGGTGCCCAATC 240
QY 484 AGACGGGGGAGGGGAGGCTCTCCATATATAGCCATCTGAATCTGCGAAGCTGACC 543
DB 241 AGACAGGGGAGCTAGCCACATCTCCAAACACAAAGTCACTGAGCTATTGTGAACAGACT 300
QY 544 TGAGGGTCACGTCGGTGGTTTACGGGTTTAAACAACCTTCCGGACAGATTTAAAGACTTC 603
DB 301 CGACGGTCATGACAGCTGTGACAGATTTAATATCTGCGGACAGATTTAAAGACTTC 360
QY 604 TGCTGTATTGAGATGCGGCAATATTACTGCTTTATAGATACCGCGGATAAGTGTCAA 663
DB 361 TCTTGTATTGAGATGCGGCAATATTACTGCTTTATAGATCAACCGAAGAAATGTCAA 420
QY 664 AGAACCCTTCTTGTGCTGGGATTAAGTCCCTCACTCCACATTTGCCAGAGGCAAG 723
DB 421 AGAACCCTTCTTGTGCTGGGATTAAGTCCCTCACTCCACATTTGCCAGAGGCAAG 480
QY 724 CAATCCGGGAATCTTGGGCGCAAGAAACAGCAGGGAACCAACGGTGTGCGAGTCT 783
DB 481 CAATTCGGGAGTCTTGGGCGCGAGAAACCAACGATAGGGAACACAGACTAGAGGTCT 540
QY 784 TCTGCTGGGCGAGACACCCCGAGAGGACACACCCCGGACCTTTCAGATATGCTGAAT 843
DB 541 TCTGCTGGGCGAGACACCCCGAGAGGACACACCCCGGACCTTTCGACATGCTTAAGT 600
QY 844 TTGAGAGTCAGACAGCACCAGACATCTTATGTGGAATACAGACACTTCTTCACT 903
DB 601 TTGAGAGTCAGACAGCACCAGGACATCTTATGTGGAATATAGACACATCTTCAACC 660
QY 904 TGTCTCTGAAGGAAGTGTGTTTCTCAGGTGGTAAAGTACTTCTGCGCCGACACTGAGT 963
DB 661 TGTCTCTGAAGGAAGTGTGTTTCTTAGTGGGTGAGCACTTCTGTCACACCCAGAGT 720

QY 964 TTGTTTTCACGGCGGATGACAGATGTTTTTGTGAACACCCATCATCATCTCTGAATTACTGA 1023
DB 721 TTGTTTTCACGGCGGATGACAGTGTGTTGTGAACACCCATCATCATCTCTGAATTACTGA 780
QY 1024 ATAGTTTATCCAAAGACCAAGCCAAAGATCTCTTCATAGGTGATGTGATCCCAATGCTG 1083
DB 781 ATAGCTTATCCAAAGACCAAGCCAAAGACTTGTTCATAGGTGATGTGATCCCAATGCTG 840
QY 1084 GACCTCATCGGGATGAAGAAGCTGAAGTACTACATCCCAGAAAGTTGTTTACTCTGGGCTCT 1143
DB 841 GGCTTCACCGGATGAAGAAGCTGAAGTACTACATCCCAGAAAGTCTCTACACCGGCTCT 900
QY 1144 ACCACCCCTATCCAGGGGAGGGGGTCTCTACTCCGGCCACTGCGCTGAGGCTGT 1203
DB 901 ACCACCCCTATCCAGGGGAGGGGGTGTGATTTCTCTACTCCGGCCCTTGCCTTGAGGCTGT 960
QY 1204 ACCATATCATCTGACGAGCTCCATCTCTACCCCATTTGATGAGCTTTATATATGGAATGSCC 1263
DB 961 ACAGTGGCAGTACGGGGTCCATCTCTACCCCTATTGATGATGTTTATACGGGAATGSCC 1020
QY 1264 TTCAGAAACTCGGCTCGTTTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATCCAGG 1323
DB 1021 TTCAGAAACTCGGCTCGTTTCCAGAGAAAGCACAAGGCTTCAGGACATTTGATATGAAG 1080
QY 1324 AGAAAAACAAAAATAACATCTGCTCTATGTAGATCTGATGTTAGTACATAGTAAAAAC 1383
DB 1081 AGAAAAATAAGAAAAATATTGTTCTCTATATAGACCTAATGTTAGTACATAGCAGAAAAAC 1140
QY 1384 CPCAAGAGATGATGATATTTGGTCTCAGTTCAGAGTTCAGAGTGTCTATTTAAATGTAA 1439
DB 1141 CTCAGAGATGATGATATATATGTTCTGCTCAGTTCAGAGTTCCTAAATTTAAATGTCTGA 1196

RESULT 14

AA167865
ID AA167865 standard; cDNA; 1196 BP.
XX
AC AA167865;
XX
DI 13-MAR-2002 (first entry)
XX
DE Murine LIG46 polypeptide encoding cDNA.
XX
KW Leptin; LIG46; LIG56; leptin induced gene; Tgtp; LRG-47; RC10-II; Stral3;
XX anorectic; anabolic; antisense therapy; mouse; ss.
XX Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 3..1194
FT /tag= a
FT /product= "LIG46"
FT /note= "the coding sequence is specifically claimed"
FT sig_peptide 3..98
FT mat_peptide /tag= b
FT 99..1193
FT /tag= c
XX
XX US2001024808-A1.
XX
XX 27-SEP-2001.
XX
XX 12-MAR-2001; 2001US-0804357.
XX
XX 29-OCT-1998; 98US-106378P.
XX 19-NOV-1998; 98US-D195896.
XX 10-SEP-1998; 98US-0150857.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX White D, Zhou J, Tartaglia LA;
XX

DR WPI: 2001-624963/72.
 DR P-PSDB: AAG66115, AAG66116.
 XX An isolated polypeptide useful for treating a weight disorder and for
 PT screening compounds that may inhibit protein activity comprises a
 PT leptin-induced protein
 XX Claim 16; Fig 1; 46pp; English.
 XX The invention relates to genes whose expression are induced by leptin.
 CC LIG46 and LIG56 are novel leptin induced genes (LIG), while four other
 CC genes such as LRP-47, RC10-11 and Stral3 have been previously
 CC identified. The leptin induced proteins can be expressed by standard
 CC recombinant methodology. The proteins and encoding polynucleotides may
 CC be used in screening assays to identify compounds that may bind to it.
 CC Administering a molecule, e.g., an antisense molecule, which reduces
 CC expression of activity of protein selected from LIG46, LIG56, LRP-47,
 CC LRP-47, RC10-11, and Stral3 may be used to treat a weight disorder. The
 CC method may further comprise administering leptin. The present sequence
 CC represents a cDNA encoding a murine LIG46 protein.
 XX Sequence 1196 BP; 350 A; 269 C; 290 G; 287 T; 0 other;

Query Match 53.1%; Score 906.4; DB 23; Length 1196;
 Best Local Similarity 84.9%; Pred. No. 1.2e-237;
 Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 244 AATGAGTGTGGACGTCGAGAGATAAAGTTGTGGTATCTCTGATGGCAATGCT 303
 DB 1 AGATGAGTGTGGGCGGTGGAAGATCAAGTTGCTGGGCATCTCTGATGGCAATGCT 50
 QY 304 TCATTTATTTATATCGAAGTCTCCAAAGCAGTAGCCAAAGAAATGGAAGGGG 363
 DB 61 TCATTTATTTGAGTCTCCAAAGCAGTAGCCAAAGAAATGGAAGGGG 120
 QY 364 AAGTAATAATACCAAGAGAGTTCTGGAAGATATCTACCCCTCCGAGGCACTATGGA 423
 DB 121 GAGTAATAATCCGAAAGAGAGTTCTGGAAGCAGCCAGCAGCTCCCGGGCATCTGGA 180
 QY 424 ACCGAGAGCAGAGCAAGCTGAACCGGCGAGTACAGCCCACTCTGAGCAGTCCGACCAAC 483
 DB 181 ACAGGAGCAGAGCAAGCTGAACAGGTTGACAGTCCCACTCTGAGCAGTCCGACCAAC 240
 QY 484 AGACGGGGGAGCGGCGAGGCTCTCCAAATATAGCCATCTGAACCTGCGAACCCTGACC 543
 DB 241 AGACGGGGAGCTAGCCACATCTCCAAACACAGTCACCTGAGCTATTGTGACACAGACT 300
 QY 544 TGAGGGTCAGTCTGCTGGTGTACGGGTTTAAACACCTTGGCGGACAGATTAAAGACTTTC 603
 DB 301 CGACGGTCATGACAGCTGTGACAGATTTTAATAATCTGCGGACAGATTAAAGACTTTC 360
 QY 604 TGCTGTATTTGAGATCGCGCAATTAITCAGTCTGCTTATAGATCAGCCGAGTAAGTGCCAA 663
 DB 361 TCTGTATTTGAGATCGCGCAATTAITCAGTCTGCTTATAGATCAGCCGAGTAAGTGCCAA 420
 QY 664 AGAAACCTTTCTTTCTGCGGATTAAGTCCCTCTACCTCCACATTTTGGCAGAGCAAG 723
 DB 421 AGAAGCCCTCTTACTATTGGCGATAAAGTCCCTCTATTCACATTTTGGCAGAGCAAG 480
 QY 724 CAATCCGGGATCTCTGGGCGCAAGAGCAAGCAGGAGCAACCGTGGTGGTGGTGGTCT 783
 DB 481 CAATCCGGGATCTCTGGGCGGAGAAACCAAGTGGGAGCAACCGTGGTGGTGGTGGTCT 540
 QY 784 TCCTGCTGGGCGACACACCCCGAGGAGCAACACCCCGACCTTTTCAGATATGCTGAAT 843
 DB 541 TCCTGCTGGGCGACACACCCCGAGGAGCAACACCCCGACCTTTTCAGATATGCTGAAT 600
 QY 844 TTGAGAGTGAGAGCAGCAGACAGATTCTTATGTGGAAGTACAGAGACACTTTCTCAACT 903
 DB 601 TTGAGAGTGAGAGCAGCAGACAGATTCTTATGTGGAAGTACAGAGACACTTTCTCAACT 660
 QY 904 TGTCTCTGAAGAGTCTGTTCTCAGGTGGGTAACTCTCTGCCAGACACTGAGT 963
 DB 904 TGTCTCTGAAGAGTCTGTTCTCAGGTGGGTAACTCTCTGCCAGACACTGAGT 963

DB 661 TGCCCTGAAGGAAGTCTGTTCTTAGTGGGTGAGCACITCCTCTCCAGACCGAGAGT 720
 QY 954 TTGTTTTCAGGGCGATGACGATGTTTGTGAACACCCCATCACATCCTGAATTAATTGA 1023
 DB 721 TTGTTTTCAGGGCGATGACGATGTTTGTGAACACCCCATCACATCCTTAATTACTGA 780
 QY 1024 ATAGTTTATCCAGACCAAGCCAAAGATCTTTTCATAGTGTGATGATCCCAATGCTG 1083
 DB 781 ATAGCTTATCCAGAGCAAGCCAAAGACTTGTTCATAGTGTGATGATCCCAATGCTG 840
 QY 1084 GACCTCATCGGGATTAAGAAGCTGAAGTACTACATCCCAAGAGTGTGTTACTCTGGCCTCT 1143
 DB 841 GGCCTCACGGGATTAAGAAGCTGAAGTACTACATCCCAAGAGTCTTCTACACCGGCTCT 900
 QY 1144 ACCCACCTTATGACGGGGAGGGGGTTCCTTACTCCGGCCACCTGGGCCCTGAGGCTGT 1203
 DB 901 ACCCACCTTATGCGGGGGTGGTGGATTCCTTACTCCGGCCCTTGGCTTGGGCTGT 960
 QY 1204 ACCATATCACTGACGAGTCCATCTCTACCCCATTTGATGACGTTTATCTGGAATGCTGC 1263
 DB 961 ACAGTGGGACTAGCGGGTCTCTCTACCTATTGATGATGTTTATACGGGATGCTGC 1020
 QY 1264 TTCAGAACTCGGCTCTGTTCCAGAGAAACACAAAGGCTTCAGACATTTGATATCGAGG 1323
 DB 1021 TTCAGAACTCGGCTCTGTTCCAGAGAAACACAAAGGCTTCAGACATTTGATATGAGG 1080
 QY 1324 AGAAACCAAAATTAACATCTGCTCCTATGATGATGATGATGATGATGATGATGATGAT 1383
 DB 1081 AGAAACCAAAATTAACATCTGCTCCTATGATGATGATGATGATGATGATGATGATGAT 1140
 QY 1384 CTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1439
 DB 1141 CTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1196

RESULT 15
 AAD46672
 ID AAD46672 standard; cDNA; 1196 BP.
 XX AC AAD46672;
 XX DT 27-JAN-2003 (first entry)
 XX DE Murine LIG46 cDNA.
 XX KW LIG46; leptin; low body weight; chromosomal mapping; tissue typing;
 KW forensic biology; transgenic; gene therapy; antianorectic; murine;
 KW gene; ss.
 OS Mus sp.
 XX FH Key Location/Qualifiers
 FT CDS 3..1196
 FT /tag= a /product= "Murine LIG46 protein"
 FT sig_peptide 3..98
 FT /tag= b
 FT mat_peptide 99..1193
 FT /tag= c
 FT /product= "Murine mature LIG46 protein"
 XX WO200274905-A2.
 XX PD 26-SEP-2002.
 XX PF 20-NOV-2001; 2001WO-US43345.
 XX PR 21-NOV-2000; 2000US-0717778.
 XX PA (MILL-) MILLENNIUM PHARM INC.
 XX PI White DW, Zhou J, Tartaglia LA, Stricker-krongrad A, Clausen H;

DR WPI; 2002-759886/82.
DR P-PSDB; AAE29092.
XX
PT New isolated nucleic acid molecules, designated as leptin-induced genes
PT 46 (LIG46), useful for treating a subject having a disorder
PT characterized by undesirable level of LIG46 expression or activity,
PT such as low body weight -
XX
PS Claim 15; Fig 1; 90pp; English.
XX
CC The invention relates to LIG46, a gene whose expression is induced by
CC leptin. LIG46 DNA and protein are useful in treating a subject having
CC a disorder characterised by undesirable level of LIG46 expression or
CC activity, such as low body weight. They are also useful in a screening
CC assay, chromosomal mapping, tissue typing and forensic biology. The
CC probes based on the LIG46 nucleotide sequence are useful for detecting
CC transcripts or genomic sequences encoding the same or related proteins.
CC The LIG46 polypeptides are useful as immunogens for raising anti-LIG46
CC antibodies. The host cells are useful for producing non-human transgenic
CC animals. LIG46 DNA is used in gene therapy. The present sequence is
CC murine LIG46 cDNA.
XX
SQ Sequence 1196 BP; 350 A; 269 C; 290 G; 287 T; 0 other;

Query Match 53.1%; Score 906.4; DB 24; Length 1196;
Best Local Similarity 84.9%; Pred. No. 1.2e-237;
Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 244 AATGAGTGTGGACGTCGGAAGATAAAGTTGTGGGTATCTCGATGATGGCAATGCT 303
DB 1 AGATGAGTGTGGGCGTCGGAAGAGTCAAGTTGCTGGGCATCTCGATGATGGCAATGCT 60

QY 304 TCATTTATTTATTTATGGAAGTCTCCAAAGAGTAGCCAAAGAAAAATGAAAAGGGG 363
DB 61 TCATTTATTTGTTGGAAGTCTCCAAARACAGTAGCCAAARATGGAAAGGGAG 120

QY 364 AAGTAATANTACCAAGAGAGAGTCTGGAAGATATCTACCCCTCCGAGGAGTACTGA 423
DB 121 GAGTAATANTCCCGAAAGAGAGTCTGGAAGCAGCCAGCACTCCCGGGGCACTACTGA 180

QY 424 ACCGAGAGCAAGAGAGCTCAACCGGCGAGTACAAACCCCTCTCGAGCATGCTGACCAACC 483
DB 181 ACAGGGANACAGGAGAGCTGAACAGGTGGTACAACTCCATCTTGAACAGGTGGCAATC 240

QY 484 AGACGGGGAGGGGGGCGGCTCTCCAATATAAGCCATCTGAACACTACTCGCAACCTGACC 543
DB 241 AGACAGGGGAGCTAGCCACATCTCCAAACACAAAGTCACTGAGCTATTGTGAACAGACT 300

QY 544 TGAGGGTCACGTCGGTGTACGGGTTTAAACACTTGGCGGACAGATTAAAGACTTTC 603
DB 301 CGAGCGTCATGACAGCTGTGACAGATTTTAAATCTCGCGGACAGATTAAAGACTTTC 360

QY 604 TGCTGTATTTGAGATGCGGCAATTAATCTAGTCTGATATAGATACGCGGATAGTGTGCAA 663
DB 361 TCTGTATTTGAGATGCGGCAATTAATCTAGTCTGATATAGATCAACCGAAGAAATGTCAA 420

QY 664 AGAANCTTCTTCTGCTGGCGATTAAGTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 723
DB 421 AGAAGCCCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480

QY 724 CAATCCGGGAATCTGGGCGCAAGAGCAACCGGAGCAACCAACCGGTGGTGGCGAGTCT 783
DB 481 CAATTCGGGAGCTTGGGCGCGAGAACCAACCGTAGGGAACACAGAGTGTGAGGCTCT 540

QY 784 TCTGTGTGGGCGGAGACACCCCGAGAGGACCAACCCCGACCTTTCAGATATGCTGAAAT 843
DB 541 TCCTGTTGGGCGGAGACACCCCGAGAGGACCAACCCCGACCTTTCGGGACATGCTTAAGT 600

QY 844 TTGAGAGTGAGAGCAACCAAGACATCTTATGTGGAACTACAGAGACTTCTTCAACT 903
DB 601 TTGAGAGTGAGAGCAACCAAGACATCTTATGTGGAACTATAGAGACATCTTCTCAACC 660

QY 904 TGTCTCTGAAGGAGTGTCTGTTTCTCAGGTGGGTAAAGTACTTCTCTCCCGAGACACTGAGT 963

DB 661 TGTCCCTGAAGAGTGTCTTCTTAGGTGGTGAGCACCTTCTCTCCAGACGAGAGT 720
QY 964 TTGTTTCAAGGCGGATGACGATGTTTTTTGTGAACACCCATCACATCTCTGAATTAATTGA 1023
DB 721 TTGTTTCAAGGCGGATGACGATGTTTTTTGTGAACACCCATCACATCTCTTAITACTTGA 780
QY 1024 ATAGTTTATCCAAAGACCAAGCAAGATCTCTTCATAGGTGATGTGATCCACCAATGCTG 1083
DB 781 ATAGCTTATCCAAAGACCAAGCAAGACTTCTTCATAGGTGACCTGATCCACAATGCTG 840
QY 1084 GACCTCATCGGGATACAAAGCTGAAGTACTACATCCAGAGAGTGTGTTTACTCTGCGCTCT 1143
DB 841 GGCTCTACCGGGATACAAAGCTGAAGTACTACATCCAGAGAGTCTTCTACACCGCGCTCT 900
QY 1144 ACCCACCTATGAGGGGGAGGGGTTCTCTCTACTCCGGCCACCTGGCCCTGAGGCTGT 1203
DB 901 ACCCACCTATGAGGGGGAGGGTGGATTCTCTACTCCGGCCCTTGGCTTGGAGCTGT 960
QY 1204 ACCATATCACTGACAGGTCCATCTCTACCCATTTGATGACGCTTTTATCTGGAATGTGCC 1263
DB 961 ACAGTGGCACTAGCCGGGTCCATCTCTACCCCTATTGATGATGTTTTATACGGGAATGTGCC 1020
QY 1264 TTCAGAAACTCGGCTCTGTTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATCGAGG 1323
DB 1021 TTCAGAAACTCGGCTCTGTTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATTTGAAG 1080
QY 1324 AGAAACAAACAAATACATCTGCTCTATGTAGATCTGATTTAGTACATAGTAGAANAAC 1383
DB 1081 AGAAATTAAGAAAAATATTTGTTCTATATAGACCTAATGTGTAGTACATAGCAGANAAC 1140
QY 1384 CTCAGAGAGATGATTTGATATTTGGTCTCAGTTGCGAGAGTGTCTCATTTAAATGCTAA 1439
DB 1141 CTCAGAGAGATGATTTGATATCTGCTCTCAGTTGCGAAGTCTCTAATTTAAATGCTCA 1196

Search completed: October 20, 2003, 14:46:01
Job time : 488.67 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2003, 14:27:23 : Search time 3857.36 Seconds
(without alignments)
10755.472 Million cell updates/sec

Title: US-09-804-357b-13

Perfect score: 1707

Sequence: 1 acgcgtccgcgcagcggcag.....aatctgttgatggccctt 1707

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	899.2	52.7	1201	9	AL553132
2	717.2	42.0	832	9	AU137528
3	708.2	41.5	876	13	BX370083
4	659.6	38.6	816	12	B1546809

5	556.8	38.5	743	14	CB554366
6	651.2	38.1	692	10	BG106243
7	648.6	38.0	1053	12	BM557603
8	624.4	36.6	725	12	B1549804
9	573.2	33.6	1432	12	BM926933
10	565	33.1	659	10	BG741357
11	530	31.0	632	10	BG721682
12	508.6	29.8	680	14	CB457941
13	488.2	28.6	701	9	AV724042
14	463.6	27.2	608	10	BF001654
15	458.8	26.9	593	10	AW958548
16	433.2	25.4	558	12	BM537435
17	431.4	25.3	857	14	W26453
18	429.4	25.2	547	14	CD560301
19	422.4	24.7	945	13	B0369620
20	414.4	24.3	587	10	BE623172
21	410	24.0	801	10	BE622082
22	408.6	23.9	609	10	BE100018
23	407.8	23.9	436	9	AV732823
24	406.6	23.8	589	9	AW495241
25	404.4	23.7	1084	10	BE894128
26	399.6	23.4	592	10	BF467111
27	398.8	23.4	424	9	AV682596
28	397	23.3	865	13	B0427178
29	395.4	23.2	686	14	BY754595
30	393.6	23.1	505	9	AA204363
31	390.8	22.9	664	14	BY724171
32	385.4	22.6	655	14	BY753386
33	379.4	22.2	517	9	AA119132
34	370.2	21.7	488	10	BE632125
35	366.8	21.5	476	9	AA986646
36	364.8	21.4	457	9	AA733556
37	358.2	21.0	543	9	AW530344
38	345	20.2	610	10	BB619419
39	341	20.0	964	13	BQ893660
40	335.6	19.7	501	10	BE685687
41	334.8	19.6	490	9	A1408668
42	334.2	19.6	489	10	BG062994
43	328.2	18.2	480	14	H78875
44	326.8	19.1	410	10	BF469505
45	326.8	19.1	484	9	A1407754

ALIGNMENTS

RESULT 1
AL553132
LOCUS
DEFINITION AL553132 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA clone CS0DI074YK14 5-PRIME, mRNA sequence.
ACCESSION AL553132
VERSION AL553132.2
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 15, 2001 this sequence version replaced gi:12892683.
Contact: Genoscope
Genoscope - Centre National de Sequençage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 1713.f For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DI074F07QPl&cluster=1713.f. Contact : Feng Liang Email : fliang@lifetech.com URL :

CB554366	MMSP0063-
BG106243	602289589
BM557603	AGENCOURT
B1549804	603194750
BM926933	AGENCOURT
BG741357	602631909
BG721682	602695830
CB457941	7116060 MA
AV724042	AV724042
BF001654	7g91c06.x
AW958548	EST370618
BM537435	ha82C09.g
W26453	30a6 Human
CD560301	B0410D02-
B0369620	603596654
BE623172	up49a10.x
BE622082	601440945
BE100018	UI-R-BJ1-
AV732823	AV732823
AW495241	UI-M-BH3-
BE894128	601438257
BF467111	UI-M-CG0P
AV682596	AV682596
B0427178	603955596
BY754595	BY754595
AA204363	mu30ell.r
BY724171	BY724171
BY753386	BY753386
AA119132	mp66f04.r
BE632125	uu57d09.y
AA986646	uc82d09.y
AA733556	vu7lf09.r
AW530344	UI-R-BU0-
BB619419	BB619419
BQ893660	AGENCOURT
BE685687	uu57d09.x
A1408668	EST236959
BG062994	H3002H04-
H78875	yu09g06.r1
BF469505	UI-M-BH3-
A1407754	EST236044

```
http://fulllength.invitrogen.com/ Invitrogen Corporation 160C
Faraday Avenue Genoscope sequence ID : CS0D1074BF07QPI.
FEATURES
    source
        1..1201
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0D1074YK14"
            /tissue_type="PLACENTA COT 25-NORMALIZED"
            /note="1st strand cDNA was primed with a NotI-oligo(dT)
            primer. Five prime end enriched, double-strand cDNA was
            digested with Not I and cloned into the Not I and EcoR V
            sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      302 a 280 c 331 g 251 t 37 others
ORIGIN
    Query Match      52.7%; Score 899.2; DB 9; Length 1201;
    Best Local Similarity 98.5%; Pred. No. 3.7e-196;
    Matches 954; Conservative 5; Mismatches 4; Indels 6; Gaps 5;
Qy 9 GCGAGCGGAGCGGAGCGGACAGCGGCAACAAAGTCCGGAGGCTAGCAGACCCAGCGGAG 68
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 80 GCGACGCGNAGCGGAGCGGAGCGGCAACAAAGTCCGGA-GTAGCAGACCCAGCGGAG 138
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 69 CAGTCCCTGCGCGGACACCGCGCGCCGCTCCGCGGCGCGCGCATGGAGCGTGAG 128
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 139 CAGTCCCTGCGCGGACACCGCGCGCCGCTCCGCGGCGCGCGCATGGAGCGTGAG 198
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 129 CTCGGCGGTGCGCGGCTAGCGCGGAGCGCGGAGCGCGGAGCGTGAGTGTGCGCGGATC 188
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 199 CTGCGGCGGTGCGCGGCTAGCGCGGCGGA--GGCGGAGCGTGAGTGTGCGCGGATC 256
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 189 TCCGCGCTTGCCCGCGCGCGGAGCTGGAGCTGCTCCGCGACAGATATGAGAAATG 248
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 257 TCCGCGCTTGCCCGCGCGCGGAGCTGGAGCTGCTCCGCGACAGATATGAGAAATG 316
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 249 AGTGTGGAGCTGCAAGAAATTAAGTGTGGGTATCTGATGATGCAAAATGCTTCATT 308
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 317 AGTGTGGAGCTGCAAGAAATTAAGTGTGGGTATCTGATGATGCAAAATGCTTCATT 376
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 309 TATTTTATTTAGGAAGTCTCCAAAGCAGTAGGCAAGAAATAATGGAAAGGGGAAGTA 368
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 377 TATTTTATTTAGGAAGTCTCCAAAGCAGTAGGCAAGAAATAATGGAAAGGGGAAGTA 436
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 369 ATAATACCAAGAGAAGTCTCGAAGATATCTACCCCTCCGAGGCATCTGGAACCA 428
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 437 ATAATACCAAGAGAAGTCTCGAAGATATCTACCCCTCCGAGGCATCTGGAACCA 496
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 429 GAGCAAGAGAAGCTGAACCGGCGAGTACACCCCATCTGAGCATGCTGACCAACAGAG 488
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 497 GAGCAAGAGAAGCTGAACCGGCGAGTACACCCCATCTGAGCATGCTGACCAACAGAG 556
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 489 GGGAGGCGGCGAGGCTCCATATATAGCCATCTGAAGTCTGCAAGCTGACCTGAG 548
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 557 GGGAGGCGGCGAGGCTCCATATATAGCCATCTGAAGTCTGCAAGCTGACCTGAG 616
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 549 GTCAGTCCGTGTTAGCGGTTTAAACAGTTCGCCGAGAGATTTAAAGACTTTCCTG 608
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 617 GTCAGTCCGTGTTAGCGGTTTAAACAGTTCGCCGAGAGATTTAAAGACTTTCCTG 676
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 609 TATTTGAGATCGCGCAATTTACCTGCTTATAGATCAGCCGATTAAGTGTGCAAGAA 568
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 677 TATTTGAGATCGCGCAATTTACCTGCTTATAGATCAGCCGATTAAGTGTGCAAGAA 736
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 669 CTTTCTTGTGTTGGGATTAAGTCCCTCACTCCACATTTTCCGAGAGGCAAGCAATC 728
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 737 CTTTCTTGTGTTGGGATTAAGTCCCTCACTCCACATTTTCCGAGAGGCAAGCAATC 796
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 729 CGGAATCTCTGGGCCAAGAAAGCAACGCGAGGAACCAACGCTGGTGGAGTCTTCCTG 788
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 797 CGGAATCTCTGGGCCAAGAAAGCAACGCGAGGAACCAACGCTGGTGGAGTCTTCCTG 855
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Qy 789 CTGGCCAGACACCCCCAGAGGACACCCAGCGACCTTTCAGATATGCTGAATTTGAG 848
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 856 CTGGCCAGACACCCCCAGAGGACACCCAGCGACCTTTCAGATATGCTGAATTTGAG 915
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 849 AGTGAGAAGACCAAGACATTTTATGTGGAACCTACAGAGACACTTTCTTCAACTTGTCT 908
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 916 AGTGAGAAGACCAAGACATTTTATGTGGGATACAGAGACACTTTCTTCAACTTGTCT 975
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 909 CTGAAGGAAG-TGCTGTTTCTCAGTGGGTAAAGTACTTCTGCCAGACACTGAGTTGT 967
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 976 CTGAAGGAAGTGTGTTTCTCAGTGGGTAAAGTACTT-CTGCCAGACACTGAGTTGT 1034
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 968 TTTCAGGG 976
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1035 TTTCAGGS 1043

RESULT 2
AUI37528
LOCUS AUI37528 832 bp mRNA linear EST 02-AUG-2002
DEFINITION AUI37528 PLACE1 Homo sapiens cDNA clone PLACE1006678 5', mRNA
sequence.
ACCESSION AUI37528
VERSION AUI37528.1 GI:10998067
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 832)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
    source
        Location/Qualifiers
            1..832
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="PLACE1006678"
                /tissue_type="placenta"
                /clone_lib="PLACE1"
                /note="vector: pME18SFL3"
BASE COUNT 248 a 193 c 194 g 194 t 3 others
ORIGIN
    Query Match      42.0%; Score 717.2; DB 9; Length 832;
    Best Local Similarity 96.2%; Pred. No. 2.8e-154;
    Matches 787; Conservative 0; Mismatches 25; Indels 6; Gaps 5;
Qy 320 GGAAGTCTCCAAAAGCAGTAGCCCAAGAAAAAATGAAAAAGGGGAAGTAATATACCCAA 379
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 9 GAAAGTCTCCAAAAGCAGTAGCCCAAGAAAAAATGAAAAAGGGGAAGTAATATACCCAA 68
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 380 AGAGAAGTCTGGAAGATATCTACCCCTCCGAGGCATCTGGAACCGAGAGCAAGAA 439
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 69 AGAGAAGTCTGGAAGATATCTACCCCTCCGAGGCATCTGGAACCGAGAGCAAGAA 128
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 440 GCTGAACCGGCGAGTACACCCCTCTTGAGCATCTGACCAACAGACGCGGAGCGGG 499
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
129 GCTGAACCGGCGAGTACACCCCTCTTGAGCATCTGACCAACAGACGCGGAGCGGG 188
```


FEATURES	source	Location/Qualifiers
1. .725		
		/organism="Homo sapiens"
		/mol_type="mRNA"
		/db_xref="taxon:9606"
		/clone="IMAGE:5274205"
		/tissue_type="hippocampus"
		/lab_host="DH10B"
		/clone_lib="NIH_MGC_95"
		/note="Organ: brain; Vector: pBluescriptR {modified pBluescript KS+; Site.1: BamHI; Site.2: SalI-xhoI (glucyag 5'-oligo-dn primed using primer 5'-TTT:TTTTTTT:TTTTVN-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."}
BASE COUNT	183 a	198 c 216 g 128 t
ORIGIN		
Query Match	36.6%	Score 624.4; DB 12; Length 725;
Best Local Similarity	96.9%	Pred. No. 6.5e-133;
Matches	650;	Conservative 0; Mismatches 16; Indels 6; Gaps 3;
QY	9	GGCGAGCGGAGCGGCGAGCAGCGGCAACAAGTGTCCGGAGGCTAGCAGAGCCCAAGCCGGAG 68
Db	14	GGCGAGCGGAGCGGCGAGCAGCGGCAACAAGTGTCCGGAGGCTAGCAGAGCCCAAGCCGGAG 73
QY	69	CAGTCCCTGCCGCGACACCGCGGGCGCCGCTCCGGGGCGCCGCGCATGGAGCTGAG 128
Db	74	CAGTCCCTGCCGCGACACCGCGGGCGCCGCTCCGGGGCGCCGCGCATGGAGCTGAG 133
QY	129	CTCGCGCGGTGCCGGGCTGAGCGCGCGGAGCGCGGCGAGCTGGATGTGGCCCGGATC 188
Db	134	CTCGCGCGGTGCCGGGCTGAGCGCGCGGAGCGCGGCGAGCTGGATGTGGCCCGGATC 193
QY	189	TCCGCGCCTTGCCTCCGCGCGCGGAGCTGGAGCTGCTCCGGACAAGATATGAGAAATG 248
Db	194	TCCGCGCCTTGCCTCCGCGCGCGGAGCTGGAGCTGCTCCGGACAAGATATGAGAAATG 253
QY	249	AGTGTGGAGCTGCAAGATAAAGCTTGTGGGTATCTGTGATGATGCAATGTCTTCATT 308
Db	254	AGTGTGGAGCTGCAAGATAAAGCTTGTGGGTATCTGTGATGATGCAATGTCTTCATT 313
QY	309	TATTTTATTGGAAGTCTCCAAAGCAGTAGCCAGAGAAAATAAGTGGAAAAGGGGAAATA 368
Db	314	TATTTTATTGGAAGTCTCCAAAGCAGTAGCCAGAGAAAATAAGTGGAAAAGGGGAAATA 373
QY	369	ATAATACCAAGAGAGAGTCTTGGAGATATCTACCCCTCCGAGGATCTACTGGAACCGA 426
Db	374	ATAATACCAAGAGAGAGTCTTGGAGATATCTACCCCTCCGAGGATCTACTGGAACCGA 433
QY	429	GAGCAAGAGAGCTGAAACGGCAGTACAACCCCATCTGAGCATGCTGACCAACCAAGCAG 488
Db	434	GAGCAAGAGAGCTGAAACGGCAGTACAACCCCATCTGAGCATGCTGACCAACCAAGCAG 493
QY	489	GGGAGGGGGGCGAGCTCTCCAATATAGCGCATCTGAACCTACTGCGAACCTGACCTGAG 548
Db	494	GGGAGGGGGGCGAGCTCTCCAATATAGCGCATCTGAACCTACTGCGAACCTGACCTGAG 553
QY	549	GTCACTCGGTG-GTTAGGGGTTTTAACAA-CTTGGCGGACAGATTAAAGACATTCTTCG 606
Db	554	GTCACTCGGACGTTAAGGGGTTTTACAACCTTGGCGGACAGATTAAAGACATTCTTCG 613
QY	607	TG--TATTTGAGATCCGCAAA-TTATTCAGTGGTTATAGATCAAGCCGGATA-AGTGTGCA 662
Db	614	TGTTACTCCGAGATCCGCAAACTTATTCAGTGGTTATAGATCAAGCCGGATAAGTGTGCA 673
QY	663	AGAAAACCTTTCTTGTGCTGGGATTAAGTCCCTCACTCCACATTTTGCCA 714
Db	674	AGAAAACCTTTCTTGTGCTGGGATTAAGTCCCTCACTCCACATTTTAGCA 725

RESULT 9	BM926933	1432 bp	mRNA	linear	EST 12-MAR-2002
LOCUS	AGENCOURT_6682000 NIH_MGC_121				Homo sapiens cDNA clone IMAGE:5767816
DEFINITION	5', mRNA sequence.				
ACCESSION	BM926933				
VERSION	BM926933.1	GI:19377312			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 1432)				
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished Contact: Robert Strausberg, Ph.D. Email: qcapbs@email.nih.gov Tissue Procurement: Life Technologies, Inc. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM12826 row: k column: 17 High quality sequence stop: 493.				
FEATURES	Location/Qualifiers				
source	1. 1432				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:5767816"				
	/lab_host="DH10B"				
	/clone.lib="NIH_MGC_121"				
	/notes="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 3 fetal brains, female age 20 weeks, female age 24 weeks, and male age 26 weeks. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 0.7-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 017. Note: this is a NIH_MGC Library."				
BASE COUNT	343 a 464 c 306 g 304 t				15 others
ORIGIN					
	Query Match 33.68; Score 573.2; DB 12; Length 1432;				
	Rest Local Similarity 97.8; Pred. No. 4.4e-12i;				
	Matches 581; Conservative 0; Mismatches 13; Indels 0; Gaps 0;				
Cy	9 GCGCAGCGGACGCGGACGCGGCAACAGTGC CGGAGGCTACGAGGCAACGCGGAG 68				
Dd	1 GCGCAGCGGACGCGGACGCGGCAACAGTGC CGGAGGCTACGAGGCAACGCGGAG 60				
Qy	69 CAGTCCCTGCGCGGACACGCGCGGCGCGCCCGTCCGGGGCGCGCGCATGAGCGGTGAG 128				
Dd	61 CAGTCCCTGCGCGGACACGCGCGGCGCGCCCGTCCGGGGCGCGCGCATGAGCGGTGAG 120				
Qy	129 CTGCGGCGGTGCGCGGCGTGA GCGCGGAGCGGCGGACGTGGATGTGGCGCGCATC 188				
Dd	121 CTGCGGCGGTGCGCGGCGTGA GCGCGGAGCGGCGGACGTGGATGTGGCGCGCATC 180				
Qy	189 TCCCGCCCTTGC CCGCCCGCCCGCGGAGCTGAGCTGCTCCCGGACAAAGATATGAGAAATG 248				
Dd	181 TCCCGCCCTTGC CCGCCCGCCCGCGGAGCTGAGCTGCTCCCGGACAAAGATATGAGAAATG 240				
Qy	249 AGTGTGGACGTCGAAGAATAAAGTTGTGGGTATCTCTGATGATGGCAAAATGCTTTCATT 308				
Dd	241 AGTGTGGACGTCGAAGAATAAAGTTGTGGGTATCTCTGATGATGGCAAAATGCTTTCATT 300				

```

QY 309 TATTTTATTATGAGTCTCCAAAGCAGTAGCCAAAGAAAAAATGGAAGAGGGAAGTA 368
DB 301 IATTTTATTATGAGTCTCCAAAGCAGTAGCCAAAGAAAAAATGGAAGAGGGAAGTA 360
CY 369 ATAATACCCAAAGAGAGTCTTGGAGATATCTACCCCTCCCGAGGCATCTGGAACCGA 428
DB 361 ATAATACCCAAAGAGAGTCTTGGAGATATCTACCCCTCCCGAGGCATCTGGAACCGA 420
QY 429 GAGCAAGAGAGTGTGAACCGGAGTCAACCCCATCTGAGCATGCTGACCAACAGAGG 488
DB 421 GAGCAAGAGAGTGTGAACCGGAGTCAACCCCATCTGAGCATGCTGACCAACAGAGG 480
QY 489 GGGGAGCGGCGAGGCTCTCCATATAAGCCATCTGAACCTACTGCAACCTGACCTGAGG 548
DB 481 GGGGAGCGGCGAGGCTCTCCATATAAGCCATCTGAACCTACTGCAACCTGACCTGAGG 540
QY 549 GTCACGTGGTGGTTACAGGTTTAAACAATTCGCGGACAGATTTAAAGACTTT 602
DB 541 GTCCCGTGGTGGTACAGGTTTAAACAATTCGCGGATGATTTGAAGAGTTT 594

RESULT 10
BG741357
LOCUS 602631909P1 NCI_CGAP_Skn3 659 bp mRNA linear EST 15-MAY-2001
DEFINITION 602631909P1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4776866 5',
mRNA sequence.
ACCESSION BG741357
VERSION BG741357.1 GI:14052010
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10630 row: b column: 03
High quality sequence stop: 659.
Location/Qualifiers
FEATURES
source
1..659
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4776866"
/lab_host="DH10B (T1 phage-resistant)"
/notice="Organ: skin; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: oligo dr.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 173 a 168 c 155 g 163 t
ORIGIN
Query Match 33.1%; Score 565; DB 10; Length 659;
Best Local Similarity 96.1%; Pred. No. 3e-119;
Matches 635; Conservative 0; Mismatches 15; Indels 11; Gaps 5;
QY 586 ACAGATTTAAAGACTTTCTGCTGATTGAGATCCCGCAATTTATCTACTGCTTATGATC 645
DB 1 ACAGATTTAAAGACTTTCTGCTGATTGAGATCCCGCAATTTATCTACTGCTTATGATC 60
QY 646 AGCCGGATAAGTGTGCAAGAAACCTTTCTGTGTGGCGCATTAAGTCCCTCACCTCCAC 705

```

```

DB 61 AGCCGGATAAGTGTGCAAGAAACCTTTCTTG-TGCTGGGATTAAAGTCCCTCACTCCAC 119
QY 706 ATTTTGCAGAGGCAAGCAATCCGGGAATCTCTGGGGCCCAAGAAACCAACGCGAGGAACC 765
DB 120 ATTTTGCAGAGGCAAGCAATCCGGGAATCTCTGGGGCCCAAGAAACCAACGCGAGGAACC 179
QY 766 AAAGGTGGTGGCGAGTCTTCTGCTGGCCAGACACCCCGAGAGGACAAACCCCGGACC 825
DB 180 AAAGGTGGTGGCGAGTCTTCTGCTGGCCAGACACCCCGAGAGGACAAACCCCGGACC 239
QY 826 TTTTCAAGATATGCTGAAATTTGAGAGTGAAGAGCACCAGACATTTCTTATGTGGAATACA 885
DB 240 TTTTCAAGATATGCTGAAATTTGAGAGTGAAGAGCACCAGACATTTCTTATGTGGAATACA 299
QY 886 GAGACACTTTTCAACTTGTCTGCTGAAGGAAGTGTGTTTC-----TCAGGTGGGTAAG 940
DB 300 GAGACACTTTTCAACTTGTCTGCTGAAGGAAGTGTGTTTCCTTTCAGGTGGGTAAG 359
QY 941 TACTTCC-TGCCCCAGACACTGAGTTGTTTTC---AAGGGCGATGACGATGTTTTTGTGA 996
DB 360 TACTTCCGTGCCAGACACGTGAGTTGTTTCTCAAGGGCGGATGACGATGTTTTTGTGA 419
QY 997 ACACCATACATCCTGAATTTACTTGAATAGTTTATCCAGAGCACCAGCAAGCAAGATCTCT 1056
DB 420 ACACCATACATCCTGAATTTACTTGAATAGTTTATCCAGAGCACCAGCAAGATCTCT 479
QY 1057 TCATAGGTGATGATCCACAATGCTGGACCTCATCGGATAAGAAAGCTCAACTACTACA 1116
DB 480 TCATAGGTGATGATCCACAATGCTGGACCTCATCGGATAAGAAAGCTCAACTACTACA 539
QY 1117 TCCAGAGATTTTACTTCTGGCTCTACCCACCTTATGCAGGGGAGGGGGTTCCTCT 1176
DB 540 TCCAGAGATTTTACTTCTGGCTCTA-CCACCTATGCAGGGGAGGGGGTTCCTCT 598
QY 1177 ACTCGGGCACCTGGCCCTGAGGCTGACCATATCACTGACGAGTCCATCTCTACCCCA 1236
DB 599 ACTCGGGCACCTGGCCCTGAGGCTGACCATATCACTGACGAGTCCATCTCTACCCCA 658
QY 1237 T 1237
DB 659 T 659

RESULT 11
BG721682
LOCUS 602695830P1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:482158 5',
DEFINITION mRNA sequence.
ACCESSION BG721682
VERSION BG721682
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 632)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cagpbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10745 row: k column: 07
High quality sequence stop: 625.
Location/Qualifiers
FEATURES
source
1..632

```

FEATURES
SOURCE

```

/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 6BOV"
/notes="vector: pcDNA3.1; Site_1: EcoRI; Site_2: NotI;
Library made with RNA pooled from multiple tissues
including liver, lung, hypothalamus, pituitary, and
placenta/endorium."
a 149 c 165 q 158 t
208 a

```

D _b	355	TCTCAATTACTGTGAACCTGCACCTGAGGGTCATGTTCAGTGGTTTCAGGTTTCACAGCTTT	414
Q _y	581	GCCGGACAGATTTAAAGACATTTCTGCTGTATTTTGAGATGCCGCAATATTTCACCTGCTTAT	640
D _b	415	GCCAGACAGATTTAAAGATTTCTGCTATATTTCAGATGTGCAAAATTATTTCACCTGCTTAT	474
Q _y	641	AGATCAGCCGGATAAGTGTGCAAGAAACCTTTCTTGTCTGGCGATTAAATCCCTCAC	700
D _b	475	AGACCAACAGATAAGTGTCAAGAAAGCCCTCTTATTGCTGGCATTTAAGTCCCTAAC	534
Q _y	701	TCACATTTTGCAGAGGCAAGCAATCCGGGATCCTGGGGCCAAAGCAACGCAGG	760
D _b	535	TTCACATTTTCGATAGAGGCAAGCAATTCGGGAATCTTGGGGAAGAAACCAACCATGTGGG	594

```

QY 761 GAACAAACGGTGGTGGAGTCTTCTGCTGGCCAGACACCCCGAGAGGACAAACCC 820
D 761 GAACAAACGGTGGTGGAGTCTTCTGCTGGCCAGACACCCCGAGAGGACAAACCC 820
D 595 GAACAAACAGTGGTGGAGTCTTCTGCTGGCCAGACACCCCGAGAGGACAAACCC 654
QY 821 CGACCTTTCAGAI 833
D 655 TGACCTTTCAGAT 667

RESULT 13
AV724042
LOCUS AV724042 HTB Homo sapiens cDNA clone HTBAXG01 5', mRNA linear EST 16-OCT-2000
DEFINITION AV724042 HTB Homo sapiens cDNA clone HTBAXG01 5', mRNA linear EST 16-OCT-2000
ACCESSION AV724042
VERSION AV724042.1 GI:10828039
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 701)
Gu.Y., Peng.Y., Song.H., Huang.Q., Yang.Y., Gao.G., Xiao.H., Xu.X.,
Li.N., Qian.B., Liu.F., Qu.J., Gao.X., Cheng.Z., Xu.Z., Zeng.L., Xu
.S., Gu.W., Tu.Y., Jia.J., Fu.G., Ren.S., Zhong.M., Lu.G., Hu.R.,
Chen.J., Chen.Z. and Han.Z.
HTB Homo sapiens cDNA HTB clones
Unpublished
Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzq@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1..701
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HTBAXG01"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="HTB"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 183 a 159 c 177 g 177 t 5 others
ORIGIN
Query Match 28.6%; Score 488.2; DB 9; Length 701;
Best Local Similarity 93.2%; Pred. No. 1.5e-101;
Matches 564; Conservative 0; Mismatches 32; Indels 9; Gaps 5;

QY 489 GGGAGGGGGGCGAGCTCTCCATATAGCCATCTGAAGTACTGCGAAGCTGACCTGAGG 548
D 103 GGAGAACAGAGAGAGCTCTCCATATAGCCATCTGAAGTACTGCGAAGCTGACCTGAGG 162
QY 549 GTCACGTCGGTGTACGGGTTTAAACAACTTCCCGGACAGATTTAAAGACATTTCTGCTG 608
D 163 GTCACGTCGGTGTACGGGTTTAAACAACTTCCCGGACAGATTTAAAGACATTTCTGCTG 222
QY 609 TATTGAGATGCCCAATTATTCAGTCTTATAGATCAGCGGATAGTGTGCAAGAAA 668
D 223 TATTGAGATGCCCAATTATTCAGTCTTATAGATCAGCGGATAGTGTGCAAGAAA 282
QY 669 CCTTCTGTGTGCTGGCGATTAACTCCCTCAGTCCCAATTTGCCAGAGGCAAGCAATC 728
D 283 CCTTCTGTGTGCTGGCGATTAACTCCCTCAGTCCCAATTTGCCAGAGGCAAGCAATC 342
QY 729 CGGGAATCCTGGGCGCCAGAAAGCAACGCGAGGGAACCAACGGTGTGCGAGTCTTCCTG 788
D 343 CGGGAATCCTGGGCGCCAGAAAGCAACGCGAGGGAACCAACGGTGTGCGAGTCTTCCTG 402

```

```

QY 789 CTGGGCGAGACACCCCGAGAGGACAAACCCCGAGGACCTTTTCAGATATGCTGAAATTTGAG 848
D 403 CTGGGCGAGACACCCCGAGAGGACAAACCCCGAGGACCTTTTCAGATATGCTGAAATTTGAG 462
QY 849 AGTCAGAGCACCACAGACATTTCTATGTGGAATCAGACAGACACTTCTTCAACTTGTCT 908
D 463 AGTCAGAGCACCACAGACATTTCTATGTGGAATCAGACAGACACTTCTTCAACTTGTCT 522
QY 909 CTGAAGGAAGTGTGTTTCTCAGTGGGTAAGTACTTCTGCCCGACACACTGAGTTTGT 968
D 523 CTGAAGGAAGTGTGTTTCTCAGTGGGTAAGTACTTCTGCCCGACACACTGAGTTTGT 582
QY 969 TTCAAGGGCGATGACGATGTTTTTGTGGAACACCCATCCTGATTTACATTTGATAGT 1028
D 583 TTCAAGGGCGATGACGATGTTTTTGTGGAACACCCATCCTGATTTACATTTGATAGT 639
QY 1029 TTATCCAAAGACCAAGCCAAAGATCTTTCATAGGTGATGTCACACACTGCTGACCT 1088
D 640 TTATCCAAAGACCAAGCCAAAGATCTTTCATAGGTGATGTCACACACTGCTGACCT 693
QY 1089 CATCG 1093
D 694 ATTGG 698

RESULT 14
BF001654/c
LOCUS BF001654
DEFINITION 7991C06.x1 NCI-CGAP_Col6 Homo sapiens cDNA clone IMAGE:3313834 3'
Similar to TR:Q9Y2A9 Q9Y2A9 TYPE II MEMBRANE PROTEIN. ;, mRNA
sequence.
ACCESSION BF001654
VERSION BF001654.1 GI:10701929
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 608)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.
, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL, send email to:
info@image.lnl.gov
Seq primer: -40up from Gibco
High quality sequence stop: 465.
FEATURES
source
1..608
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3313834"
/tissue_type="Colon tumor, RER+"
/lab_host="DH10B"
/clone_lib="NCI-CGAP_Col6"
/notes="Organ: colon; Vector: pT73D-pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI-CGAP_Col6 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1057416-1061255, and 1144584-1145351)."

```



```
BASE COUNT      181 a   128 c   121 g   178 t
ORIGIN
Query Match      27.2%; Score 463.6; DB 10; Length 608;
Best Local Similarity 91.6%; Pred. No. 6.5e-96;
Matches 522; Conservative 8; Mismatches 33; Indels 7; Gaps 4;
QY 1068 GTATCCACAATGCTGGACCTCATCGGCATAGCAAGCTGAAGTACTACATCCAGAGIT 1127
DB 608 GTATCCACAATGCTGGACCTCATCGGTGAAGTAGCTGAAGTACTACATCTCAGTAGIT 549
QY 1128 GTTACTCGGCTCTACCCACCTATGCGAGGGAGGGGTTCTCTACTCCGGCCAC 1187
DB 548 GTTACTCGGCTCTACACACATATGCGAGGTGTAGGAGGTTCTCTACTCCGGCCAC 489
QY 1188 CTGGCCCTGAGGTGTACCATATACACTGACAGGTGCATCTTACCCCAATTGAAGC 1247
DB 488 CTGGCCCTGAGGTGTACCATATACACTGACAGGTGCATCTTACCCCAATTGAAGC 429
QY 1248 TATACTGGAATGTCCTTCAGAACTCGGCCTGTTCCAGAGAAACAAAGGCTTCAGG 1307
DB 428 TATACTGGAATGTCCTTCAGAACTCGGCCTGTTCCAGAGAAACAAAGGCTTCAGG 369
QY 1308 ACATTTGATATCGAGGAGAAAACAAAATAAATCTGCTCTATGTAGATCTGATGTTA 1367
DB 368 ACATTTGATATCGAGGAGAAAACAAAATAAATCTGCTCTATGTAGATCTGATGTTA 309
QY 1368 GTACATAGTAGAAACCTCAAGAGATGATTGATATTTGGTCTCAGTTGCGAGTGCAT 1427
DB 308 GTACATAGTAGAAACCTCAAGAGATGATTGATATTTGGTCTCAGTTGCGAGTGCAT 249
QY 1428 TTAAGTCTAAATATAGATCAAACTCAATTTGSGATGAAAGGGTWTTTTGRATGGY 1487
DB 248 TTAAGTCTAAATATAGATCAAACTCAATTTGSGATGAAAGGGTWTTTTGRATGGY 189
QY 1488 CCCATGTTGGGCTCACATTAGATTAATTCTATTNAAC--ATGAAATTCGCTTAT 1545
DB 188 CCCATGTTGGTCTCACATAGAGTAATTCATATAACCATGAATATGCCITAT 129
QY 1546 GAGTGATACCCATTT-ANGGCTCTAANCCTT---CATTTGNACTACGTGAAGAGGGA 1601
DB 128 GAGTGATACCCATTTGAGGCTCTAAACCTTCAATTTGCTACTCAGTGAAG-AGGGA 70
QY 1602 AAGCGGAGAGGTAATTINTTATGCTGA 1631
DB 69 AAGCGGAGAGTGGTAATTTCTTTATGGA 40

RESULT 15
AW958548      593 bp  mRNA  Linear  EST 01-JUN-2000
LOCUS      AW958548
DEFINITION  EST370618 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION  AW958548
VERSION    AW958548.1 GI:8148232
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            J.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
            Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
            Unpublished
            Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
```

```
Plate: 121
Seq primer: Reverse.
            Location/Qualifiers
            source
            l..593
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone_lib="MAGE resequences, MAGE"
            /note="Vector: pBluescriptSKm"
BASE COUNT      127 a   112 c   114 g   109 t   131 others
ORIGIN
Query Match      26.9%; Score 458.8; DB 10; Length 593;
Best Local Similarity 99.6%; Pred. No. 8.4e-95;
Matches 460; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 708 TTGGCCAGGAAGCAAGCAATCCGGGNAATCCTGGGCCAAGAAAGCAACCGAGGAACCAA 767
DB 1 TTGGCCAGGAAGCAAGCAATCCGGGNAATCCTGGGCCAAGAAAGCAACCGAGGAACCAA 60
QY 768 ACGGTGGTGGGAGTCTTCTGCTGGGCCAGACACCCCCAGAGACACACCCCGACCTT 827
DB 61 ACGGTGGTGGGAGTCTTCTGCTGGGCCAGACACCCCCAGAGACACACCCCGACCTT 120
QY 828 TCAGATATGCTGAAATTTGAGAGTGAGAAGCACCACCAAGACATCTTATGTGGAACACAGA 887
DB 121 TCAGATATGCTGAAATTTGAGAGTGAGAAGCACCACCAAGACATCTTATGTGGAACACAGA 180
QY 888 GACACTTCTTCAACTTGTCTCTGAAGGAAGTGCTGTTTCTCAGGTGGGTAAGTACTTCC 947
DB 181 GACACTTCTTCAACTTGTCTCTGAAGGAAGTGCTGTTTCTCAGGTGGGTAAGTACTTCC 240
QY 948 TGCCCGAGACACTGAGTTGTTTTCAGAGGCGATGACGATGTTTTCGTAACACCCCATCAC 1007
DB 241 TGCCCGAGACACTGAGTTGTTTTCAGAGGCGGATGACGATGTTTTCGTAACACCCCATCAC 300
QY 1008 ATCCTGAATTACTTGAATAGTTTATCCAGACCAAGCCCAAGATCTCTTCATAGTGAT 1067
DB 301 ATCCTGAATTACTTGAATAGTTTATCCAGACCAAGCCCAAGATCTCTTCATAGTGAT 360
QY 1068 GTATCCACAATGCTGGACCTCATCGGGATAAGAACTGAAGTACTACATCCAGAGATT 1127
DB 361 GTATCCACAATGCTGGACCTCATCGGGATAAGAAAGCTGAAGTACTACATCCAGAGATT 420
QY 1128 GTTACTCTGGCCTCTACCCACCTATGCGAGGGGAGGGGG 1169
DB 421 GTTACTCTGGCCTCTACCCACCTATGCGAGGGGAGGGGG 462
```

Search completed: October 20, 2003, 19:37:17
Job time : 3863.36 secs

Db 419 CNTGGGMMGCTGGGNGNTGG-----CNMNGNMGNCARYNNAARYTNGTNT 469
Qy 796 AGACACCCAGAGAGACAAACACCCGACCTTTTCAGATATGCTGAAATTTGACAGTGAGA 855
Db 470 TTYTNYTNGNGTNGCNGMNSGNCNCNCNGCNCARYTNYTNGCNYTNGCNYTNGCNY 529
Qy 856 AGCACAAGACATCTTATGTTGGAACACAGACACTTTTCTCAACTTGTCTCTCAAGG 915
Db 530 ARTTYGAYGAYATHVTCARTGGAYTYACNGARGAYTYTYAAYTNNACNYTNAARG 589
Qy 916 AAGTCTCTTTCTCAGGTGGTGAAGTACTTCTGCGCCAGACACTGAGTTGTTTCAAGG 975
Db 590 AYTNYCAAYTTCARMTNGTNGTNGCNGTGTCCNCARGNCAYTTTATGTITNAARG 649
Qy 976 GCGATGACGATGTTTGTGTGAACACCATCACATCTCTGAATTTACTTGAATGATTTATCCA 1035
Db 650 GNGAYGAYGAYTNTYTNCAYTNCNAAAYTNYTNGARTTYTNGAYGGT----- 703
Qy 1036 AGACCAAGCCAAAGATCTCTTCATAGTGATGTGATCCACAATGCTGGACCTCATCGG 1095
Db 704 GGGAYCCGNCARGAYTNTYNTNGTNGNGAYGTNATHMNCARGCNYTCCNAAATMGNA 763
Qy 1096 ATAAGAAGCTGAATGACTACATCCAGAAAGTTGTTTAC---TCTGGCCTTACCCACCT 1152
Db 764 AYACNAARGTNAARTAYTYATHCCNCNCNSNATGTAYMNGMNCNACNAYTAYCCNCT 823
Qy 1153 ATGAGGGGAGGGGGTTCCTCTACTCCGCCACCTGGCCCTGAGGCTGTACCATATCA 1212
Db 824 AYCGMNGMNGGNGNTAYTNTATGSMNGMNCNAGTNGMNGMNYTNCARGCNATHA 883
Qy 1213 CTGACAGGTCATCTCTACCCATTCATGACGTTTATACGGAATGTGCTTCAGAAAC 1272
Db 884 TGGARGAYCCNARYTNYTNSNATHGAYGAYTNTTYTNGNGATGTGYTNGMNGNY 943
Qy 1273 TCGGCTCGTCCAGAGAACAAAGCTTCAGACATTTGATATC---GAGGAGAAA 1329
Db 944 TGGNFTNMSNCNATGAYCAYCGNGNTTYAARCNTTYGGNATHMNGMNCNYTNG 1003
Qy 1330 AAAAAATAACATCTGCTCTCTATGATCTGATGTTAGTATAGTAGAATAACCTCAAG 1389
Db 1004 AYCCNYTNGAYCCNTGYTNTAYMNGNGNYTNYTNGINGINCMYMGNYTMSNCCNYTNG 1063
Qy 1390 AGATGATGATATTTGGTCTCAGTTCAGTTCAGAGTGTCTCATTTAAATGC 1436
Db 1064 ARATGTGACNATGTGGGCNYTNGTNCNGAYGARGGNYTNAATGY 1110

RESULT 2

US-09-459-133-14
; Sequence 14, Application US/09459133
; Patent No. 6416988
; GENERAL INFORMATION:
; APPLICANT: Konklin, Darrell C.
; APPLICANT: Yamamoto, Gayle
; APPLICANT: Jaspers, Stephen R.
; APPLICANT: Gao, Zeren
; TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS
; FILE REFERENCE: 98-77
; CURRENT APPLICATION NUMBER: US/09/459,133
; CURRENT FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: 60/111,697
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 1167
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: degenerate sequence
; FEATURE:
; NAME/KEY: misc_feature

; LOCATION: (1)...(1167)
; OTHER INFORMATION: n = A,T,C or G
US-09-459-133-14
Query Match 9.08; Score 154.4; DB 4; Length 1167;
Best Local Similarity 33.74; Pred. No. 9e-31;
Matches 259; Conservative 109; Mismatches 382; Indels 18; Gaps 2;
Qy 669 CCTTCTGTTGCTGGCGATTAAAGTCCCTCACATCCACATTTTCCAGAAAGCAAGCAATC 728
Db 418 CCNTAYTNYTNYTNGCNGTNAARWSNGARCCNGNCAYTYTCCNGCNGMNCARGCNGTN 477
Qy 729 CGGAATCTCTGGGCCAAGAAAGCAACGAGGAAACCAACAGGTGTGGGAGTCTCTCCTG 788
Db 478 MNGARACNTGGGGNW-----SNCNGTNGCNGNACNMGNTNTNTTY 522
Qy 789 CTGGCCAGACACACCCCGAGAGCAACCCCGACCTTTCAGATATGCTGAAATTTGAG 848
Db 523 YTYTNGMNSNCCNYTNGGNATGGGNGCNGAYTNGMNSNYTNGTNCNTGGGAR 582
Qy 849 AGTGAAGACACCAAGACATCTTATGTGGAACCTCAGACACACTTCTTCAACTGTCT 908
Db 583 WSNMNGNTAYGNGGAYTNYTNTGGAAYTYTNGAYTNGCNCNTAYAYMGNACN 642
Qy 909 CTGAAGGAAGTGTCTTCTCAGGTGGTAAAGTACTTCTGCCCAGACACTGAGTTGTT 968
Db 643 YTNARGAYTNYTNYTNACTNGYTWNMCAYCAVGYCCNGAYGTNAAYTYTGN 702
Qy 969 TTCAGGGCGATGACGATGTTTTGTGAACACCATCATCATCTCTGAATTTACTTGAATG 1028
Db 703 YTCARGTNCARGAYGAYGNTTYTNGTNCAYTATCCGNCNYTNYTNGARCAAYTNCARCN 762
Qy 1029 TTATCCAGACCAACCAAGATCTCTCATAGTGTGATGATCCACAATGCTGGACCT 1088
Db 763 YTNCCNCCNACNTGGCNGMNSNYTNTAYTNGNGARATHTYACNARGCNAARCN 822
Qy 1089 CATCGGATAAGAGCTGAAGTACTATCATCCAGAGAGTGTGTTTACTCTGGCTCTACCA 1148
Db 823 YTNMGNARCCNGGNGGCCNTTYTAYTNCNAAACNTTYTYTGARGGNGAYTAYCN 882
Qy 1149 CCCTATGAGGGGAGGGGTTCTCTACTCGGCCACCTGCGCTGAGGCTGTACCA 1208
Db 883 GCNTAYGNCNWSGNGGNGNTAYGTNATHWSGNGMNYTNGCNCNTGGTYNTNCAR 942
Qy 1209 ATCACTGACAGTCTCATCTCTACCCCATTTGATGAGCTTTATATCTGGAATGTGCTCAG 1268
Db 943 GCNGCNGMNGTNGCNCNTTYCCNTTYGAYGAYTNTAYACNGNTTYTGYTNGN 1002
Qy 1269 AAACCTGCGCTCGTTCAGAGAAACAAAGGCTTCAGGACATTTGATATCGAGGAGAA 1328
Db 1003 GCNTYNGNYTNGCNCNMGNCNAYCCNGNTTYTNCAC---NGCNTGGCCNGCNGAR 1059
Qy 1329 AACAAAAATAACATCTGCTCTCTATGATCTGATGTTAGTATAGTAAAGCAACCTCAA 1388
Db 1060 MGNACMNGAYCCNTGYGNGTNGMNGNYTNYTNGTNCAYCCNGTMSNCCNAR 1119
Qy 1389 GAGATGATGATATTTGGTCTCAGTTCAGATGCGAGTGTCTCATTTAAATGC 1436
Db 1120 GAYACNATHTGGYNTGGMGNCAYTNTGGTNGCNGARYTNCARTGY 1167

RESULT 3

US-09-459-133-3
; Sequence 3, Application US/09459133
; Patent No. 6416988
; GENERAL INFORMATION:
; APPLICANT: Konklin, Darrell C.
; APPLICANT: Yamamoto, Gayle
; APPLICANT: Jaspers, Stephen R.
; APPLICANT: Gao, Zeren
; TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS
; FILE REFERENCE: 98-77
; CURRENT APPLICATION NUMBER: US/09/459,133

Query Match	8.7%	Score 148;	DB 4;	Length 1420;
Best Local Similarity	52.7%	Pred. No. 4.7e-26;		
Matches 425;	Conservative 0;	Mismatches 360;	Indels 21;	Gaps 4;
Qy	657	TGTCAAAGAAACCTTTCTTTGTTGCTGGCGATTAAAGTCCCTCACTCCACATTTTGCAGA	716	
Db	474	TGTTCCAAAGATACTTCTTTGCTGTCCTGGCCATCAAGTCACAGCCTGGTCAGGTGGAGCGA	533	
Qy	717	AGGCAACCAATCCGGGATCTCTGGGGCCACGAAGACACCGAGGGAACCAACGGTGGTG	775	
Db	534	CTGTGGGCTATCCGACACAGTGGG-----GCAGGGTGGGGGATGGGCTAGGGGC	584	
Qy	777	CGAGTCTTCTGTCTGGCCAGACACCCCGAGAGACAACCCACCCGACCTTTTCAGATATG	836	
Db	585	CGGCAGCTGAAGCTGTGTCTCTAGGGGTGGCAGGATCCGCTCCCGACGCCAGCTG	644	
Qy	837	CTGAAATTTGAGAGTGAAGACCAAGACATCTTATGTGGAACCTACAGAGACACTTTC	896	
Db	645	CTGSCCTATGAGACTAGGAGTTTGATGACATCTCCAGTGGGACTTCACCTGAGGACTTC	704	
Qy	897	TTCAACTTGTCTGAAGGAGTGCTGTCTTCTCAGTGGGTAACTACTTCTGCCCCAGAC	956	
Db	705	TTCAACCTGAGCTCAAGAGCTGCACCTGCAGCGCTGGTGGTGGCTGCCTGCCCCGAC	764	
Qy	957	ACTGAGTTTGTCTTCAAGGGCGATGACGATGTTTTTGTGAACACCCATCACATCCTGAAT	1016	
Db	765	GCCCATTTTCATGCTAAAGGGAGATGACGATGCTCTTTGTCCACGTCCCAACGTGTAGAG	824	
Qy	1017	TACTTGAATAGTTTATCCAGACCAAGCCAAAGATCTCTTTCATAGGTGATGTGATCCAC	1076	
Db	825	TTCTTGATG-CTGGGACCCAGCCAGGACCTCTCGTGGGAGATGTATCCGC	878	
Qy	1077	AATGCTGGACCTCATCGGGATAAGAAAGCTGAAGTACTACATCCCAAGATTTGTTTAC---	1133	
Db	879	CAAGCCCTGCCAACAGGAACACTAAGGTCAAAATCTTCATCCCAACCTCAATGTACAGG	938	
Qy	1134	TCTGGCCTCTACCAACCTATGACGGGGGAGGGGGTTCCTCTACTCGGCCACCTGGCC	1193	
Db	939	GCCACCACCTACCCACCTATGCTGGTGGGGGAGATATGTCATGTCAGAGCCACAGTG	998	
Qy	1194	CTGAGGCTGACCATATCTGACACAGGTCCATCTCTACCCCATGTATGACGTTTACT	1253	
Db	999	CGGCGCTCCAGGCTATCATGGAAGATGCTGAACCTCTCCATGTATGATGTCCTTTGTG	1058	

Qy	1088	TCATCGGGATAAGAGCIGAGTACTACATCCGACAGAGTTGTTTACTCTCGCCTCTACCC	1147
Db	928	GCCTCCGAACGCCGAGGAGCCCTTCTATGTGCGCAAGACCTTCTTTGAAGGGGACTATCC	987
Qy	1148	ACCTATTCAGGGGAGGGGGTTCCTCTACTCCGGCCACCTGSCCCTCAGGCTGTACCA	1207
Db	988	AGCCTATTCGAGTGGAGTGGCTATGTAATCTCAGGACGGCTGCACCCCTGGCTGTGCA	1047
Qy	1208	TATCACTGACCAAGCTCCATCTCTACCCCAITGATGACGTTTATTACTGGAATGCGCTTCA	1267
Db	1048	GGCGGCAGCTCGGTGTGCACCCCTTCCCCCTTTGATGATGTCTACACTGSCCTTCTGCTTCG	1107
Qy	1268	GAAACTCGGCCTCGTTCCAGAGAAACAAGAGCTTCAGGACCTTTCATATCGAGGAGAA	1327
Db	1108	TGCCCTGGGCTTAGCACCCCGTGCCCATCCAGGCTTCCCTACA---GCCTGGCCAGAGA	1164
Qy	1328	AAACAAAAATACATCTGCTCCTTAITGATATCTGATGTTAGTACATAGTAGNAACCTCA	1387
Db	1165	ACGTACAGGGACCCCTCGCCCGTGGAGGCGCTGCTCTTGTTGCATCCAGTCAGCCCTCA	1224
Qy	1388	AGAGATGATTGATATTTGGTCTCAGITTCAGAGTGCCTCAITTTAAATAGCTAA	1439
Db	1325	GGACACCAATTTGGCTCTGGAGACATCTGTGGGTCCTCCAGAGCTCCAGCTGGCTGA	1276

RESULT 7

```

US-09-996-243-208
: Sequence 208, Application US/09996243
: Patent No. 6478825
: GENERAL INFORMATION:
: APPLICANT: Ashkenazi, Avi J.
: APPLICANT: Baker, Kevin P.
: APPLICANT: Botstein, David
: APPLICANT: Desnoyers, Luc
: APPLICANT: Eaton, Dan L.
: APPLICANT: Ferrara, Napoleone
: APPLICANT: Fong, Sherman
: APPLICANT: Gerber, Hanspeter
: APPLICANT: Gerritsen, Mary E.
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Grimaldi, J. Christopher
: APPLICANT: Gurney, Austin L.
: APPLICANT: Kijavlin, Ivar J.
: APPLICANT: Napier, Mary A.
: APPLICANT: Pan, James
: APPLICANT: Paoni, Nicholas F.
: APPLICANT: Roy, Margaret Ann
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tumas, Daniel
: APPLICANT: Watanabe, Colin K.
: APPLICANT: Williams, P. Mickey
: APPLICANT: Wood, William I.
: APPLICANT: Zhang, Zemin
: TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
:   ACIDS
: TITLE OF INVENTION: Acids Encoding the Same
: FILE REFERENCE: P2730P1C13
: CURRENT APPLICATION NUMBER: US/09/996,243
: CURRENT FILING DATE: 2001-11-14
: PRIOR APPLICATION NUMBER: 60/049787
: PRIOR FILING DATE: 1997-06-16
: PRIOR APPLICATION NUMBER: 60/062250
: PRIOR FILING DATE: 1997-10-17
: PRIOR APPLICATION NUMBER: 60/065186
: PRIOR FILING DATE: 1997-11-12
: PRIOR APPLICATION NUMBER: 60/065311
: PRIOR FILING DATE: 1997-11-13
: PRIOR APPLICATION NUMBER: 60/066770
: PRIOR FILING DATE: 1997-11-24
: PRIOR APPLICATION NUMBER: 60/075945
: PRIOR FILING DATE: 1998-02-25
: PRIOR APPLICATION NUMBER: 60/078910
: PRIOR FILING DATE: 1998-03-20

```

; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02

; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match 3.1%; Score 53.6; DB 4: Length 2095;
Best Local Similarity 45.7%; Pred. No. 0.00031;
Matches 308; Conservative 0; Mismatches 354; Indels 12; Gaps 3;
QY 551 TTTAAAGACTTTCTGCTGCTATTTGAGATGCCGCAATTAITTCACGTGCTTATAGATCAGCGG 650
DB 475 TTTATGAGTATGAGCCGATTTACAGCAACACTTTCTTACACACTTCGAGACATCA 534
QY 651 GATAAGTGTGCAAGAAACCTTTCTTGTGTGGGATTAAGTCCCTCCTCCTCCACATTTT 710
DB 535 AACTGCTCTCATCAAAATCCATTTCTGTGTCATTTCTGGTGACCTCCACCTTCAGATGTG 594
QY 711 GCCAGAAGGCAAGCAATCCGGGAATCCTGGGGCCCAAGAAAGCAAGCAGGGAACCAACG 770
DB 595 AAAGCCAGGCGAGCCATTTAGAGTTACTTTGGGTGAAAAAAGTCTTGGTGGGATATGAG 654
QY 771 GTGGTGGGAGTCTTCTGCTGGGCAGACACCCCGAGAGACACCCCGACCTTTCA 830
DB 655 GTTCTTACATTTTCTTATTAGGCCAGAGGCTGAAAGGAGAGACAATG--TTGGCA 711
QY 831 GATATGCTGAATTTGAGAGTGAGAACCAACACATCTTTATGTGGAACCTACAGAGAC 890
DB 712 TTGCTTTAGAGGATGAACACCTTTTATGGTGACATAATCCGACAGATTTTITAGAC 771
QY 891 ACTTTCTTCAACTGCTCTCTGAAGGAAGTGTGTTTCTCAGTGGTGAAGTACTTCTGTC 950
DB 772 ACATATATACTGACCTGCTTGAACCACTTATGGCATTCAGTGGTGGTAACTGAGTTTGC 831
QY 951 CCAGACACTGATGTTTTCAGGGCGATGACGATGTTTGTGAACACCCCATCATC 1010
DB 832 CCCAATGCCAAGTACGTAATGAAGACACACACTGATGTTTTCATCAATACTGSCAATTA 891
QY 1011 CTGAATTACTTGAATAGTTTATCCAGACCAAGCAAGCAAGATCTTTCATAGGTGATGTG 1070
DB 892 GTGAA-----GTATCTTTTAAACCTTAAACCACTCAGAGAAGTTTTTCACAGGTTCCT 945
QY 1071 ATCCCAATGCTGGACCTCATCGGGATAAGAGCTGAAGTACTATATCCCAAGATGTGT 1130
DB 946 CTAATTGATAATTATTCCTATAGAGGATTTTACCAAAAACCCATATTTCTTACCAGAG 1005
QY 1131 TACTCTGGC---CTCTACCCACCTTATGACGGGGAGGGGGTTCCTCTACTCCGGCCAC 1187
DB 1006 TATCCTTTCAAGGTGTTCCCTCCATACCTGACAGTGGTGGGTATATAATGTCAGAGAT 1065
QY 1188 CTGGCCCTGAGGCTGTACCATATCATCAGCAGGTCCCATCTTACCCCATGATGAGTT 1247
DB 1066 TTGGTGCCAAGATCTATGAATATGATGGTCAAGCAAAACCCCAAGTTTGAAGATGT 1125
QY 1248 TATACTGGAATGTG 1261
DB 1126 TATGTCGGGATCTG 1139

RESULT 8

US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS


```
1 NAME/KEY: misc_feature
2 LOCATION: 8363..11741
3 OTHER INFORMATION: /note= "N-linked glycosylation
4 OTHER INFORMATION: sites at following locations: 8471, 8653, 8732, 8843,
5 FEATURE:
6 NAME/KEY: misc_feature
7 LOCATION: 7949..8009
8 OTHER INFORMATION: /note= "Predicted transmembrane
9 OTHER INFORMATION: domain"
10 FEATURE:
11 NAME/KEY: misc_feature
12 LOCATION: 8288..8348
13 OTHER INFORMATION: /note= "Predicted transmembrane
14 OTHER INFORMATION: domain"
15 FEATURE:
16 NAME/KEY: misc_feature
17 LOCATION: 9434..9494
18 OTHER INFORMATION: /note= "Predicted transmembrane
19 OTHER INFORMATION: domain"
20 FEATURE:
21 NAME/KEY: misc_feature
22 LOCATION: 10052..10112
23 OTHER INFORMATION: /note= "Predicted transmembrane
24 OTHER INFORMATION: domain"
25 FEATURE:
26 NAME/KEY: misc_feature
27 LOCATION: 10178..10238
28 OTHER INFORMATION: /note= "Predicted transmembrane
29 OTHER INFORMATION: domain"
30 FEATURE:
31 NAME/KEY: misc_feature
32 LOCATION: 10886..10946
33 OTHER INFORMATION: /note= "Predicted transmembrane
34 OTHER INFORMATION: domain"
35 FEATURE:
36 NAME/KEY: misc_feature
37 LOCATION: 10955..11015
38 OTHER INFORMATION: /note= "Predicted transmembrane
39 OTHER INFORMATION: domain"
40 FEATURE:
41 NAME/KEY: misc_feature
42 LOCATION: 11216..11276
43 OTHER INFORMATION: /note= "Predicted transmembrane
44 OTHER INFORMATION: domain"
45 FEATURE:
46 NAME/KEY: misc_feature
47 LOCATION: 11894..11954
48 OTHER INFORMATION: /note= "Predicted transmembrane
49 OTHER INFORMATION: domain"
50 FEATURE:
51 NAME/KEY: misc_feature
52 LOCATION: 12293..12353
53 OTHER INFORMATION: /note= "Predicted transmembrane
54 OTHER INFORMATION: domain"
55 FEATURE:
56 NAME/KEY: misc_feature
57 LOCATION: 12377..12437
58 OTHER INFORMATION: /note= "Predicted transmembrane
59 OTHER INFORMATION: domain"
60 FEATURE:
61 NAME/KEY: misc_feature
62 LOCATION: 212..278
63 OTHER INFORMATION: /note= "Possible hinge sequence"
64 FEATURE:
65 NAME/KEY: misc_feature
66 LOCATION: 279
67 OTHER INFORMATION: /note= "Cleavage site"
```

```
US-08-422-582-7
Query Match          2.7%; Score 45.8; DB 4; Length 14148;
Best Local Similarity 54.5%; Pred. No. 0.089;
Matches 116; Conservative 0; Mismatches 92; Indels 5; Gaps 1;
```

Search completed: October 20, 2003, 19:40:45
Job time : 129.782 secs

```
Qy 5 GTCCGCCAGCGCGCAGCGCGGCAACAAAGTCCCGAGGCTAGCAGAGCCAGGCC 64
Db 137 GGCCCCGCTGAGCTGGGGTCTCCGCGCGCGGGCGGCGCTGGGACGCGGGCCATGCG 196
Qy 65 GGAGCAGTCC-----CTGCGCGCGACACCGCGGGCGCCGCTCCGGGGCGCGCATG 115
Db 197 CGCGCTGCCCTAACGATGCCGCGCGCGCGCGCGCGCTGGCGCTGGGCCCTG 256
Qy 120 GAGCGTGAAGTGGGGCGGTGCGCGGGGCTGAGCGCGCGCGGGAGCGCGGGGACGTGGATGTG 179
Db 257 GGCCTGTGGCTCGGGGCGCTGGGGGGGCGCGCGGGCGCGCTGCGGGCGCTGCGAGGCC 316
Qy 180 GCCGCGATCTCCCGCCCTTGGCCCCCGCGCGCC 212
Db 317 CCTGCTCTGCGGCGCCAGCGCGCGCGCGCGCC 349
```


Qy	61	AGCGGAGCAGTCCCTGCGGCCACACCGCGCGGCGCGCGTCCGGGGCGCCGCGCATGG	120
Db	61	AGCGGAGCAGTCCCTGCGGCCACACCGCGCGGCGCGCGTCCGGGGCGCGCGCATGG	120
Qy	121	AGCGTGAGCTGCGGCGTCCCGGGCTGAGCCGCGCGGAGCGCCGGGAGCTGCATGTGG	190
Db	121	AGCGTGAGCTGCGGCGTCCCGGGCTGAGCCGCGCGGAGCGCCGGGAGCTGCATGTGG	180
Qy	181	CCGGGATCCCGGCCCTTGCCCGCGGCCCGCGAGCTGGAGCTGCTCCGGGACAAATAT	240
Db	181	CCCGGATCTCCCGCCCTTGCCCGCGGCCCGAGCTGGAGCTGCTCCGGGACAAATAT	240
Qy	241	GAGAAATGAGTGTGGACGTCGAAGATAAAGTTGTTGGTATCCTGATGATGGCAATG	300
Db	241	GAGAAATGAGTGTGGACGTCGAAGATAAAGTTGTTGGTATCCTGATGATGGCAATG	300
Qy	301	TCCTCATTTATTTATATGGAAGTCTCCAAAAGCAGTAGCCAAGAAAAATATGGAAG	360
Db	301	TCCTCATTTATTTATATGGAAGTCTCCAAAAGCAGTAGCCAAGAAAAATATGGAAG	360
Qy	361	GGGAAGTAATATACCCAAAGAGAAGTTCTGGAAGATATCTACCCCTCCCGAGGCATCT	420
Db	361	GGGAGTAAATATACCCAAAGAGAAGTTCTGGAAGTATCTACCCCTCCCGAGGCATCT	420
Qy	421	GGAAACGAGAGCAAGAGACTGAACGGCGAGTACAACCCCATCTTGAGCATGCTGACCA	480
Db	421	GGAAACGAGAGCAAGAGACTGAACGGCGAGTACAACCCCATCTTGAGCATGCTGACCA	480
Qy	481	ACCAGACGGGGAGGGCGGCGAGCTCTCCAATATAAGCCATCTGAAGTCTGCGAACCTG	540
Db	481	ACCAGACGGGGAGGGCGGCGAGCTCTCCNATATAGCCATCTGAAGTCTGCGAACCTG	540
Qy	541	ACCTGAGGTCACGTGCGTGGTTCACGGTTTTAACACTINGCCGACAGATTTTAAGACT	600
Db	541	ACCTGAGGTCACGTGCGTGGTTCACGGTTTTAACACTINGCCGACAGATTTTAAGACT	600
Qy	601	TTCTGCTGATTTGAGATGCGCAATTTATTCAGTGTCTATAGATCAGCGGATGAAGTGTG	660
Db	601	TTCTGCTGATTTGAGATGCGCAATTTATTCAGTGTCTATAGATCAGCGGATGAAGTGTG	660
Qy	661	CAAGAACAACCTTCTTGTTCGCTGGGATTAAGTCCCTCACTCCACATTTTGCAGAGGC	720
Db	661	CAAGAACAACCTTCTTGTTCGCTGGGATTAAGTCCCTCACTCCACATTTTGCAGAGGC	720
Qy	721	AAGCAATCCGGGAATCCTGGGGCCAGAAAGCAACGAGGAACCAAGCGTGTGCGAG	780
Db	721	AAGCAATCCGGGAATCCTGGGGCCAGAAAGCAACGAGGAACCAACCGTGTGCGAG	780
Qy	781	TCCTTCCTGCTGGGCGAGACACCCCGAGAGACCAACCCCGACCTTTTCAGATATCTGA	840
Db	781	TCCTTCCTGCTGGGCGAGACACCCCGAGAGACCAACCCCGACCTTTTCAGATATCTGA	840
Qy	841	AATTTGAGGTGAGAAGCACCAAGACATTCCTATGTGGAATCAGAGACACCTTCTTCA	900
Db	841	AATTTGAGGTGAGAAGCACCAAGACATTCCTATGTGGAATCAGAGACACCTTCTTCA	900
Qy	901	ACTGTCTCTGAGGAAGTCTGTTTCTCAGGTGGTAACTCTCTGAGTGTGATCCAGACCTG	960
Db	901	ACTGTCTCTGAGGAAGTCTGTTTCTCAGGTGGTAACTCTCTGAGTGTGATCCAGACCTG	960
Qy	961	AGTTTGTTTTCAGGGCGATGAGATGTTTTCTGGAACCCCATCACATCCTCAATTACT	1020
Db	961	AGTTTGTTTTCAGGGCGATGAGATGTTTTCTGGAACCCCATCACATCCTCAATTACT	1020
Qy	1021	TGAATAGTTTATCCAAGACCAAAAGCCAAAGATCTCTTCATAGTGTGATGCCAAATG	1080
Db	1021	TGAATAGTTTATCCAAGACCAAAAGCCAAAGATCTCTTCATAGTGTGATGCCAAATG	1080
Qy	1081	CTGACCTCATCGGGATAGAAGCTGAAGTACTACATCCAGAAGTTGTTTACTCTGGCC	1140
Db	1081	CTGACCTCATCGGGATAGAAGCTGAAGTACTACATCCAGAAGTTGTTTACTCTGGCC	1140

Qy	1141	TCTACCCACCCCTATGACAGGGGGAGGGGGTTCCTCTACTCCGGCCACCTGGCCCTCAGGC	1200
Db	1141	TCTACCCACCCCTATGACAGGGGGAGGGGGTTCCTCTACTCCGGCCACCTGGCCCTCAGGC	1200
Qy	1201	TGTACCATATCACTGCACAGGTCCTACTCTACACCCCATTTGATGACGTTTATACITGGAAATGT	1260
Db	1201	TGTACCATATCACTGCACAGGTCCTACTCTACACCCCATTTGATGACGTTTATACITGGAAATGT	1260
Qy	1261	GCCTTCAGAAACTCGGGCTCGTTCCAGAGAAACACAAAGGCTTCAGACATTTGATATCG	1320
Db	1261	GCCTTCAGAAACTCGGGCTCGTTCCAGAGAAACACAAAGGCTTCAGACATTTGATATCG	1320
Qy	1321	AGGAGAAAAACAAAAATAACATCTGCTCCTCTATGTAGATCTGATGTTTAGTACATAGTAGAA	1380
Db	1321	AGGAGAAAAACAAAAATAACATCTGCTCCTCTATGTAGATCTGATGTTTAGTACATAGTAGAA	1380
Qy	1381	AACCTCAAGAGATGATTGATATTGGTCTCAGTTGCAGAGTGCCTATTTAAATGCTAAA	1440
Db	1381	AACCTCAAGAGATGATTGATATTGGTCTCAGTTGCAGAGTGCCTATTTAAATGCTAAA	1440
Qy	1441	ATAGATACAAACTCAATTTKGSATWRAAGGGGTWTTTGRATWGYCCCATCTTGGGGT	1500
Db	1441	ATAGATACAAACTCAATTTKGSATWRAAGGGGTWTTTGRATWGYCCCATCTTGGGGT	1500
Qy	1501	CTCACATTAGAGTAATTTCTATTNNAACATGAAATGGCTTTTATGAGTGATACCCATTT	1560
Db	1501	CTCACATTAGAGTAATTTCTATTNNAACATGAAATGGCTTTTATGAGTGATACCCATTT	1560
Qy	1561	ANGGCCCTTAACCTTCATTIGNACTCACGTGAAGAGGAAAGCGGAGAGCTAATTI	1620
Db	1561	ANGGCCCTTAACCTTCATTIGNACTCACGTGAAGAGGAAAGCGGAGAGCTAATTI	1620
Qy	1621	NTTTATGGTGAATGGCAGGATATTGGTCTGACTTACCGNTAGGGGANTTTAAACTGGNC	1680
Db	1621	NTTTATGGTGAATGGCAGGATATTGGTCTGACTTACCGNTAGGGGANTTTAAACTGGNC	1680
Qy	1681	CTTTTCAATCTGTTTGGATGGCCCTI	1707
Db	1681	CTTTTCAATCTGTTTGGATGGCCCTI	1707
RESULT 2			
US-09-804-006-13			
: Sequence 13, Application US/09804006			
: Patent No. US20020119517A1			
: GENERAL INFORMATION:			
: APPLICANT: White, David			
: APPLICANT: Zhou, Jianghong			
: APPLICANT: Tartaglia, Louis A.			
: TITLE OF INVENTION: LEPTIN INDUCED GENES			
: FILE REFERENCE: 07334/126001			
: CURRENT APPLICATION NUMBER: US/09/804.006			
: CURRENT FILING DATE: 2001-03-12			
: PRIOR APPLICATION NUMBER: US 09/292,228			
: PRIOR FILING DATE: 1999-04-15			
: PRIOR APPLICATION NUMBER: US 60/108,379			
: PRIOR FILING DATE: 1998-10-29			
: PRIOR APPLICATION NUMBER: US 09/150,857			
: PRIOR FILING DATE: 1998-09-10			
: NUMBER OF SEQ ID NOS: 17			
: SOFTWARE: FastSeq for Windows Version 3.0			
: SEQ ID NO 13			
: LENGTH: 1707			
: TYPE: DNA			
: ORGANISM: Homo sapiens			
: FEATURE:			
: NAME/KEY: CDS			
: LOCATION: (246)...(1436)			
: NAME/KEY: misc_feature			
: LOCATION: (1)...(1707)			
: OTHER INFORMATION: n = A,T,C or G			
US-09-804-006-13			

Query Match	99.3%;	Score 1694.8;	DB 10;	Length 1707;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1707;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ACGCTCCGCGAGCGCGCAGCGGCACAGCGGCAACAAGTGGCGGAGGCTACGAGAGCCA	60	
DB	1	ACGCGTCCGCGCAGCGCGCAGCGCAGCGGCAACAAGTGGCGGAGGCTACGAGAGCCA	60	
QY	61	AGCCGAGCAGTCCCTGCCGCGACACACCGCGGGCCGCTCCGGGGCGCGCGCATGG	120	
DB	61	AGCCGGAGCAGTCCCTGCCGCGACACACCGCGGGCCGCTCCGGGGCGCGCGCATGG	120	
QY	121	AGCGTAGCTGCGGGCTGCCGGGCTGAGCCGCGCGGAGCGCGCGGAGCTGGATGTGG	180	
DB	121	AGCGTAGCTGCGGGCTGCCGGGCTGAGCCGCGCGGAGCGCGCGGAGCTGGATGTGG	180	
QY	181	CGCGGATCTCCGCGCTTCCGCCCGCCCGCGAGCTGGAGTGTGGGTATCTCTGCGGACATAT	240	
DB	181	CGCGGATCTCCGCGCTTCCGCCCGCCCGCGAGCTGGAGTGTGGGTATCTCTGCGGACATAT	240	
QY	241	GAGAAATGAGTGTGACGCTCGAAGAAATAAAGTTGTGGGTATCTCTGATGATGGCAAAATG	300	
DB	241	GAGAAATGAGTGTGACGCTCGAAGAAATAAAGTTGTGGGTATCTCTGATGATGGCAAAATG	300	
QY	301	TCTTCATTTATTTTATTTATTTAGAGTCTCCAAAAGCAGTAGCCAAAGAAAAATGGAAG	360	
DB	301	TCTTCATTTATTTTATTTATTTAGAGTCTCCAAAAGCAGTAGCCAAAGAAAAATGGAAG	360	
QY	361	GGGAAGTAATATACCAAGAGAACTTCTGGAGATATCTACCCCTCCCGAGGCATACT	420	
DB	361	GGGAAGTAATATACCAAGAGAACTTCTGGAGATATCTACCCCTCCCGAGGCATACT	420	
QY	421	GGAAACGAGCAGCAAGAGAACTGAACCGGCGAGTACAACCCCATCTCTGAGCATGCTGACCA	480	
DB	421	GGAAACGAGCAGCAAGAGAACTGAACCGGCGAGTACAACCCCATCTCTGAGCATGCTGACCA	480	
QY	481	ACCAGAGGGGGAGCGGGCAGGCTCTCCAATATAGCCATCTGAACTACTGCGAACCTG	540	
DB	481	ACCAGAGGGGGAGCGGGCAGGCTCTCCAATATAGCCATCTGAACTACTGCGAACCTG	540	
QY	541	ACCTGAGGCTCAGCTCGGTGGTITACGGTITTTAACAACTTGCGGACAGATTTAAAGACT	600	
DB	541	ACCTGAGGCTCAGCTCGGTGGTITACGGTITTTAACAACTTGCGGACAGATTTAAAGACT	600	
QY	601	TCTCGCTGATTTGAGATGCCCAATATTTCATCTGCTTATAGATCAGCGGATAGTGTG	660	
DB	601	TCTCGCTGATTTGAGATGCCCAATATTTCATCTGCTTATAGATCAGCGGATAGTGTG	660	
QY	661	CAAGAAGAACCTTTCTTTGCTGGGATTAAGTCCCTCACTCCACATTTTGCCAGAGGC	720	
DB	661	CAAGAAGAACCTTTCTTTGCTGGGATTAAGTCCCTCACTCCACATTTTGCCAGAGGC	720	
QY	721	AAGCAATCCGGGAATCTTGGGGCAAGAAAGCAACGCGGGAACCAACCGTGGTGGAG	780	
DB	721	AAGCAATCCGGGAATCTTGGGGCAAGAAAGCAACGCGGGAACCAACCGTGGTGGAG	780	
QY	781	TCTTCTGCTGGCGCAGACACCCCGCAGAGCAACCAACCCCGACCTTTCAGATATGCTGA	840	
DB	781	TCTTCTGCTGGCGCAGACACCCCGCAGAGCAACCAACCCCGACCTTTCAGATATGCTGA	840	
QY	841	AATTTGAGAGTGAGAGCACCAGACATCTTATGTGGAACTACAGAGACACTTCTTCA	900	
DB	841	AATTTGAGAGTGAGAGCACCAGACATCTTATGTGGAACTACAGAGACACTTCTTCA	900	
QY	901	ACTTGTCTCTGAAGGAAGTCTGTTTCTCAGGTGGGTAACTTCTGCGCCAGACACTG	960	
DB	901	ACTTGTCTCTGAAGGAAGTCTGTTTCTCAGGTGGGTAACTTCTGCGCCAGACACTG	960	
QY	961	AGTTTGTGTTTCAAGGGCGATGACGATGTTTTTGTGAACCCCATCATCCTGAAATTA	1020	
DB	961	AGTTTGTGTTTCAAGGGCGATGACGATGTTTTTGTGAACCCCATCATCCTGAAATTA	1020	
QY	1021	TGAATAGTATTCGAAGACCAAGCAAGATCTCTTCATAGGTGATGTGATGCACCAATG	1080	

Db	1021	TGAATAGTTTATCCAAGACCAAAAGACTCTCTTCATAGGTGATGTCATCCACAATG	1080
Qy	1081	CTGGACCTCATCGGGATAGAAGCTGAAGTACTACATCCACAGAAGTGTTTACTCTGGCC	1140
Db	1081	CTGGACCTCATCGGGATAGAAGCTGAAGTACTACATCCACAGAAGTGTTTACTCTGGCC	1140
Qy	1141	TCTACCCACCTATGAGGGGGAGGGGGTTCCTCTACTCCGGCCACCTGGCCCTGAGGC	1200
Db	1141	TCTACCCACCTATGAGGGGGAGGGGGTTCCTCTACTCCGGCCACCTGGCCCTGAGGC	1200
Qy	1201	TGTACCATATCACTGACAGGTCATCTCTACCCCATTTGATGAGCTTATPACTTGGAAATG	1260
Db	1201	TGTACCATATCACTGACAGGTCATCTCTACCCCATTTGATGAGCTTATPACTTGGAAATG	1260
Qy	1261	GCCTTCAGAAACTCGGCCCTCGTTCAGAGAAAACACAAAGGCTTCAGGACATTTGATATCG	1320
Db	1261	GCCTTCAGAAACTCGGCCCTCGTTCAGAGAAAACACAAAGGCTTCAGGACATTTGATATCG	1320
Qy	1321	AGGAGAAAACAAAAATACATCTGCTCCTATGTAGATCTGATGTTAGTACATAGTAGAA	1380
Db	1321	AGGAGAAAACAAAAATACATCTGCTCCTATGTAGATCTGATGTTAGTACATAGTAGAA	1380
Qy	1381	AACTCTCAAGAGATGATGATATTGGTCTCAGTTCGACAGTGCCTCATTTAAAATGCTAAA	1440
Db	1381	AACTCTCAAGAGATGATGATATTGGTCTCAGTTCGACAGTGCCTCATTTAAAATGCTAAA	1440
Qy	1441	ATAGATACAAACTCAATTTKGSATWRAAGGGGTWTTTGRATWGYCCCATGTTGGGGT	1500
Db	1441	ATAGATACAAACTCAATTTKGSATWRAAGGGGTWTTTGRATWGYCCCATGTTGGGGT	1500
Qy	1501	CTCACATTACAGTAATTTCTATTINANCATGAAATTGCCTTTATGATGATACCCCATTT	1560
Db	1501	CTCACATTACAGTAATTTCTATTINANCATGAAATTGCCTTTATGATGATACCCCATTT	1560
Qy	1561	ANGGCCTCTAANCCTTCATTGTGNACTCACGTGAAGAAAGGGAAGCGGGAAGGTAATTT	1620
Db	1561	ANGGCCTCTAANCCTTCATTGTGNACTCACGTGAAGAAAGGGAAGCGGGAAGGTAATTT	1620
Qy	1621	NTTTATGGTGAATGGCGAGGATATTGGTCTGACTTACCGNTAGGGGANTTTAAAAC	1680
Db	1621	NTTTATGGTGAATGGCGAGGATATTGGTCTGACTTACCGNTAGGGGANTTTAAAAC	1680
Qy	1681	CTTTTGAATCTGTTTGGATGGCCCTT	1707
Db	1681	CTTTTGAATCTGTTTGGATGGCCCTT	1707
RESULT 3			
US-10-109-563-3			
; Sequence 3, Application US/10109563			
; Publication No. US20020170075A1			
; GENERAL INFORMATION:			
; APPLICANT: Leviten, Michael W.			
; APPLICANT: Phillips, Russell			
; TITLE OF INVENTION: TRANSGENIC MICE COMPRISING			
; TITLE OF INVENTION: BETA-1.3-N-ACETYLLACTOSAMINE DISRUPTIONS			
; FILE REFERENCE: R-021			
; CURRENT APPLICATION NUMBER: US/10/109,563			
; CURRENT FILING DATE: 2002-07-09			
; PRIOR APPLICATION NUMBER: US 60/280,706			
; PRIOR FILING DATE: 2001-03-29			
; NUMBER OF SEQ ID NOS: 6			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 3			
; LENGTH: 1831			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-10-109-563-3			

Query Match 89.7%; Score 1531.4; DB 13; Length 1831;
Best Local Similarity 97.5%; Pred. No. 0;
Matches 1585; Conservative 8; Mismatches 25; Indels 7;

Qy	13	AGCGCAGCGGCAGCGGCGCAACAAAGTCCCGAGGCTACCGAGGCGCAAGCCCGACCACT	72
Db	3	AGCGCAGCGGCAGCGGCGCAACAAAGTCCCGAGGCTAGCAGAGCGCAAGCCCGAGCACT	62
Qy	73	CCCTGCCGCGCAGACCGCCGGGCGCGCTCCGGGGCGCGCGCATGGAGCGTGAGCTGC	132
Db	63	CCCTGCCGCGCAGACCGCGGGGCGCGCTCCGGGGCGCGCGCATGGAGCGTGAGCTGC	122
Qy	133	GGCGGTCCCGGGGTGAGCCGCGGAGCGCGGCGGAGCGTGGATGATGGCGGGGATCTCC	192
Db	123	GGCGGTCCCGGGGTGAGCCGCGGAGCGCGGAGCGGTGGATGATGGCGGGATCTCC	182
Qy	193	GCCTTTGCCCGCCCGCGCGAGCTGGAGCTGCTCCGGACAAGATATGAGAAATGAGTG	252
Db	183	GCCTTTGCCCGCCCGCGCGAGCTGGAGCTGCTCCGGACAAGATATGAGAAATGAGTG	242
Qy	253	TTGAGCTCGAAGAATAAAGTTGTGGTATCTGATGATGGCAATGTCTTCATTAT	312
Db	243	TTGAGCTCGAAGAATAAAGTTGTGGTATCTGATGATGGCAATGTCTTCATTAT	302
Qy	313	TTATTATGGAAGTCTCCAAAGCAGTAGCCAGAAAGAAATGAAAGGGGAAGTAA	372
Db	303	TTATTATGGAAGTCTCCAAAGCAGTAGCCAGAAAGAAATGAAAGGGGAAGTAA	362
Qy	373	TACCCAAAGAGAAGTTCTCGAAGATATCTACCCCTCCGAGGCGATCTGGAACCGGAGCG	432
Db	363	TACCCAAAGAGAAGTTCTCGAAGATATCTACCCCTCCGAGGCGATCTGGAACCGGAGCG	422
Qy	433	AAGAGAAGCTGAACCGGCAGTACAACCCCATCTGAGCATGCTGACCAACGACACGGGG	492
Db	423	AAGAGAAGCTGAACCGGCAGTACAACCCCATCTGAGCATGCTGACCAACGACACGGGG	482
Qy	493	AGCGGCGAGGCTTCCAAATATAGCCATCTGAACCTACTCGAACCTGAGCTGAGGCTCA	552
Db	483	AGCGGCGAGGCTTCCAAATATAGCCATCTGAACCTACTCGAACCTGAGCTGAGGCTCA	542
Qy	553	CGTCGGTGGTTACGGGTTTTAACAACTTGGCGACAGATTAAAGACTTTCTCTGTATT	612
Db	543	CGTCGGTGGTTACGGGTTTTAACAACTTGGCGACAGATTAAAGACTTTCTCTGTATT	602
Qy	613	TGAGATGCCGAATATTTCAGTCTTATAGATCAGCGCGATAAGTGTGCAAGAAACCTT	672
Db	603	TGAGATGCCGAATATTTCAGTCTTATAGATCAGCGCGATAAGTGTGCAAGAAACCTT	662
Qy	673	TCATTGTTCTGGCGATTAAAGTCCCTCACTCCACATTTTGCAGNAGCAAGCAATCGGG	732
Db	663	TCATTGTTCTGGCGATTAAAGTCCCTCACTCCACATTTTGCAGNAGCAAGCAATCGGG	722
Qy	733	AATCCTGGGCGCAAGAAAGCAACGCGAGCAACCAACGGTGGTCCGAGTCTTCTCTCTGG	792
Db	723	AATCCTGGGCGCAAGAAAGCAACGCGAGCAACCAACGGTGGTCCGAGTCTTCTCTCTGG	782
Qy	793	GCCAGACACCCCGAGAGGACAACCAACCCCGACCTTTCAGATATGCTGAAATTTGAGAGTG	852
Db	783	GCCAGACACCCCGAGAGGACAACCAACCCCGACCTTTCAGATATGCTGAAATTTGAGAGTG	842
Qy	853	AGAAGCACCAGACATTTATGTGGAACTACAGAGACACTTCTTCAACCTGCTCTGA	912
Db	843	AGAAGCACCAGACATTTATGTGGAACTACAGAGACACTTCTTCAACCTGCTCTGA	902
Qy	913	AGGAAGTCTCTTCTCAGGTGGGTAAAGTACTTCTGCCAGACACTGAGTTTGTTTTCA	972
Db	903	AGGAAGTCTCTTCTCAGGTGGGTAAAGTACTTCTGCCAGACACTGAGTTTGTTTTCA	962
Qy	973	AGGCGATGACGATGTTTTTCTGAACACCCATCATCTCTGAATTTACTTGAATAGTTTAT	1032
Db	963	AGGCGATGACGATGTTTTTCTGAACACCCATCATCTCTGAATTTACTTGAATAGTTTAT	1022
Qy	1033	CAAGACCAAGCCAAAGACTCTTTCATAGGTGATGTATCCCAATAGCTGGACCTCATC	1092
Db	1023	CAAGACCAAGCCAAAGACTCTTTCATAGGTGATGTATCCCAATAGCTGGACCTCATC	1082

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2745 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 233..1423
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 233..328
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 329..1423
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-972-912-1

Query Match 87.2%; Score 1488; DB 10: Length 2745;
Best Local Similarity 97.3%; P-red. No. 0;
Matches 1575; Conservative 8; M-smatches 24; Indels 11; Gaps 7;

QY	21	CGGACGACGGGCAACAGTGGCGAGGCTAGCAGAGCCAGCCGAGCAGTCCCTGGCG	80
DB	10	CGGACGACGGGCAACAGTGGCGGA--CTAGCAGAGCCAGCCGAGCAGTCCCTGGCG	67
QY	81	CGGACACCGCGGGCGCGCGCGCGCGCGCGCATGGAGCGTGAGTCGGCGGGTCG	140
DB	68	CGGACACCGCGGGCGCGCGCGCGCGCGCGCATGGAGCGTGAGTCGGCGGGTCG	127
QY	141	CC-GGBCCTGAGCGCGCGGAGCGCGCGGACGTGGATGTGGCGGATCTCCGCGCCCTG	199
DB	128	CGGGGCTCAGCGCGCGCGGAGC-GCGGGACGTGGATGTGGCGGATCTCCGCGCCCTG	186
QY	200	CCCCCGCGCGGAGCTGGAGTGTCTCCGCGACAGATATGAGAAATGAGTGTGGAGC	259
DB	187	CCCCCGCGCGGAGCTGGAGTGTCTCCGCGACAGATATGAGAAATGAGTGTGGAGC	246
QY	260	TGGAAGTCTCCAAAGCAGTAGCCAGAAAAAATGGAAGGGGAGTAAATACCCAA	319
DB	247	TGGAAGTCTCCAAAGCAGTAGCCAGAAAAAATGGAAGGGGAGTAAATACCCAA	306
QY	320	GGAAGTCTCCAAAGCAGTAGCCAGAAAAAATGGAAGGGGAGTAAATACCCAA	379
DB	307	GGAAGTCTCCAAAGCAGTAGCCAGAAAAAATGGAAGGGGAGTAAATACCCAA	366
QY	380	AGAGAGTCTGGAAGATATCTACCCCTCCGAGGCATCTGGAACCGAGACAGAGNA	439
DB	367	AGAGAGTCTGGAAGATATCTACCCCTCCGAGGCATCTGGAACCGAGACAGAGNA	426
QY	440	GCTGAACCGGAGTACAAACCCATCTGAGCATGCTGACCAACCGAGCGGGAGCGGG	499
DB	427	GCTGAACCGGAGTACAAACCCATCTGAGCATGCTGACCAACCGAGCGGGAGCGGG	486
QY	500	CAGGCTCTCCAAATATAGCATCTGAATCTGCGAACCTGACCTGAGGGTCACTCGGT	559
DB	487	CAGGCTCTCCAAATATAGCATCTGAATCTGCGAACCTGACCTGAGGGTCACTCGGT	546
QY	560	GGTACGGGTTTTAAACATTCGCGACAGATTTAAAGACTTCTGCTATTTCAGATG	619
DB	547	GGTACGGGTTTTAAACATTCGCGACAGATTTAAAGACTTCTGCTATTTCAGATG	606
QY	620	CCGCAATATTACTGCTTATAGATCAGCGGATAGTGTGCAAGAAACCTTTCTTGTT	679
DB	607	CCGCAATATTACTGCTTATAGATCAGCGGATAGTGTGCAAGAAACCTTTCTTGTT	666
QY	680	GCTGGGATTAAGTCCCTCACTCCACATTTTCCAGAGGCAAGCAATCCGGGAATCCGT	739
DB	667	GCTGGGATTAAGTCCCTCACTCCACATTTTCCAGAGGCAAGCAATCCGGGAATCCGT	726

RESULT 5

US-10-109-563-1
; Sequence 1, Application US/10109563
; Publication No. US20020170075A1
; GENERAL INFORMATION:
; APPLICANT: Leviten, Michael W.
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE COMPRISING
; FILE REFERENCE: BETA-1.3-N-ACETYLLACTOSAMINE DISRUPTIONS
; CURRENT APPLICATION NUMBER: US/10/109,563

QY	740	GGGCCAAGAAGCAACCGACGGAACCAACGGTGGTCCGAGTCTTCTGCTGGGCCAGAC	799
DB	727	GGGCCAAGAAGCAACCGACGGAACCAACGGTGGTCCGAGTCTTCTGCTGGGCCAGAC	786
QY	800	ACCCCGAGAGCAACCAACCGACGCTTTCAGATATGCTGAAATTTTCAGAGTTCAGAGCA	859
DB	787	ACCCCGAGAGCAACCAACCGACGCTTTCAGATATGCTGAAATTTTCAGAGTTCAGAGCA	846
QY	860	CCAAGACATTTTATGTGGAACCTACAGACACACTTTCTTCAACTGTCTCTGAAGGAAT	919
DB	847	CCAAGACATTTTATGTGGAACCTACAGACACACTTTCTTCAACTGTCTCTGAAGGAAT	906
QY	920	GCTGTTTCTCAGTGGGTAGTACTTCTGCCACAGACACTGTTGTTTTCAGAGGCGA	979
DB	907	GCTGTTTCTCAGTGGGTAGTACTTCTGCCACAGACACTGTTGTTTTCAGAGGCGA	966
QY	980	TGAGGATGTTTTTCTGAACCCCACTCATCTCAATTTACTTTGAATAGTTTATCCAAAGC	1039
DB	967	TGAGGATGTTTTTCTGAACCCCACTCATCTCAATTTACTTTGAATAGTTTATCCAAAGC	1026
QY	1040	CAAGCCCAAGATCTCTTCATAGTGTATGATCCCAATTCCTGGAGCTCATCGGGATAA	1099
DB	1027	CAAGCCCAAGATCTCTTCATAGTGTATGATCCCAATTCCTGGAGCTCATCGGGATAA	1086
QY	1100	GAAGCTGAAGTACTACATCCAGAAAGTGTGTTTACTCTGGGCTTACCCACCTTATGAGG	1159
DB	1087	GAAGCTGAAGTACTACATCCAGAAAGTGTGTTTACTCTGGGCTTACCCACCTTATGAGG	1146
QY	1160	GGGAGGGGGTTCCTCTACTCCGCCACCTGGCCCTGAGGCTGTACCATATCACTGACCA	1219
DB	1147	GGGAGGGGGTTCCTCTACTCCGCCACCTGGCCCTGAGGCTGTACCATATCACTGACCA	1206
QY	1220	GGTCCATCTCTACCCCATTTGATGACGTTTATGTAAGTGTGCTTCAGAACTTCGCGCT	1279
DB	1207	GGTCCATCTCTACCCCATTTGATGACGTTTATGTAAGTGTGCTTCAGAACTTCGCGCT	1266
QY	1280	GGTCCAGAGAAACACAAGGCTTCAGGACATTTGATATCAGGAGAAACCAAAATAA	1339
DB	1267	GGTCCAGAGAAACACAAGGCTTCAGGACATTTGATATCAGGAGAAACCAAAATAA	1326
QY	1340	CATCTGCTCTATGATCTGATGTTAGTACATAGTAGAAACCTCAAGAGATGATTGA	1399
DB	1327	CATCTGCTCTATGATCTGATGTTAGTACATAGTAGAAACCTCAAGAGATGATTGA	1386
QY	1400	TATTTGCTCTCAGTTCAGAGTGTCTCATTTTAAATGCTTAAATAGATACAACTCAATTT	1459
DB	1387	TATTTGCTCTCAGTTCAGAGTGTCTCATTTTAAATGCTTAAATAGATACAACTCAATTT	1446
QY	1460	KGSATWGAAGGGTWTTCATGATGTCATGTCGATGTCGATGTCATGATGATGATGATG	1519
DB	1447	TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1506
QY	1520	TATTTNANC--ATGAAATTCCTTTTATGATGATGATGATGATGATGATGATGATGATG	1576
DB	1507	TATTTNANC--ATGAAATTCCTTTTATGATGATGATGATGATGATGATGATGATGATG	1566
QY	1577	---CATTTGACTCAGTGAAGAGGGAAGCGGAGAGGATGATGATGATGATGATGATG	1631
DB	1567	TCAATTTGGTACTCAGCTGAAG-AGGGAAAGCGAAGATGATGATGATGATGATGATG	1623

```
; CURRENT FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: US 60/280,706
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1260
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-109-563-1

Query Match      54.1%; Score 931.4; DB 13; Length 1260;
Best Local Similarity 84.1%; Pred. No. 6.4e-269;
Matches 1049; Conservative 1; Mismatches 197; Indels 0; Gaps 0;

QY 217 TGGAGTGTCTCCGGACACAGATATGAGAAATGAGTGTGGAGCTGCGAAGATAAAGTTGCT 276
DB 1 TGGAGTGTCTCCGGACACAGATATGAGAAATGAGTGTGGAGCTGCGAAGATAAAGTTGCT 60
QY 277 TGGGTATCTGTGATGAGGCAATGCTTTCATTTATTTATTTATGGAAGTCTCCAAAAGCA 336
DB 61 TGGGCATCTGTGATGAGGCAATGCTTTCATTTATTTATTTATGGAAGTCTCCAAAAGCA 120
QY 337 GTAGCCAGAAAAAATGGAAGGGAAGTAAATATATCCCAAGAGAGTTCTTGAAGA 396
DB 121 GTAGCCAGAAAAAATGGAAGGGAAGTAAATATATCCCAAGAGAGTTCTTGAAGC 180
QY 397 TATCTACCCCTCCGAGGATAGTGAACCGAGAGCAAGAGAGCTGAACCGGCAGTACA 456
DB 181 CACCCAGACATCCCCGGGATATCTGGAACAGGAAGCAGGAGAGTGAACAGGTGGTACA 240
QY 457 ACCCATCTGTGAGCATGCTGACCAACAGCAGCGGGGAGGGCGAGGCTCTCCAATATA 516
DB 241 ATCCCATCTGTAACAGGGTGGCCAAATCAGACAGGGGAGCTAGCCATCTCCAAACACA 300
QY 517 GCATCTGAATCTACTCGAACCTGACCTGAGGTCAGCTGGTGGTACGGGTTTAAACA 576
DB 301 GTCACCTGAGCTATTGTAACCAAGACTCGACGCTCATGACAGCTGTGACAGATTTTAATA 360
QY 577 ACTTGGCGGACACATTTAAAGACTTCTGCTGTATTGTGAGATGCCCCTAATTATTCTAC 636
DB 361 ATCTGCGGACAGATTTAAAGACTTCTGCTGTATTGTGAGATGCCCCTAATTATTCTAC 420
QY 637 TTATAGATCAGCCGGATAAGTGTCAAGAAAGACCTTCTTGTGTGGTGGATTAAGTCC 696
DB 421 TTATAGATCAACCGAAGAAATGTCAAGAAAGCCCTTCTTACTATTGGGATAAAGTCCC 480
QY 697 TCACTCCACATTTTGCAGAGGCAAGCAATCCGGAAATCCTGGGCGCCAGAGCAAGC 756
DB 481 TCATTCCACATTTTGCAGAGGCAAGCAATCCGGAGTCTTGGGCGCCAGAGCAAGC 540
QY 757 CAGGGAACCAACGGTGGTGGAGTCTTCTGCTGGGCGACACACCCCGAGAGCAAGC 816
DB 541 TAGGGAACCAACAGTACTGAGGGTCTTCTGTTGGGCAAGACACCCCGAGGCAAGC 600
QY 817 ACCCGACCTTTGAGATGCTGAAATTTGAGAGTGAGAGCAAGCAACATCTTATCT 876
DB 601 ACCCTGACCTTTGGGACATCTTAAAGTTTCAGAGTGACAAGCACCAGGACCTCTCATG 660
QY 877 GGAACCTACAGACACITTTCTTCACTGTCTGGAAGAGTGTCTTCTCAGGTGGG 936
DB 661 GGAACCTACAGACACITTTCTTCACTGTCTGGAAGAGTGTCTTCTCAGGTGGG 720
QY 937 TAACTACTTCTCCGACACATGAGTTTCTTTTCAAGGCGGATGACGATGTTTGTGA 996
DB 721 TGAGCACTTCTCTCCAGACGAGAGTTTGTCTTCAAGGCGGATGATGAGTGTCTTGTGA 780
QY 997 ACACCCATCATCTCTGAATTTACTTGAATAGTTTATCCAGACCAAGCAAGATCTCT 1056
DB 781 ACACCCATCATCTCTTAATTTACTTGAATAGCTTATCCAGAGCAAGCAAGACTTCT 840
QY 1057 TCATAGTGTATGATCCCAATTTCTGGACCTCATCGGATAGAAGTAAGTACTTACA 1116
DB 1116 TCTAGTGTATGATCCCAATTTCTGGACCTCATCGGATAGAAGTAAGTACTTACA
```

RESULT 6

```
US-09-804-357-1
; Sequence 1, Application US/09804357
; Patent No. US20010024808A1
; GENERAL INFORMATION:
; APPLICANT: Zhou, Jianhong
; APPLICANT: White, David
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/109001
; CURRENT APPLICATION NUMBER: US/09/804,357
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/195,896
; PRIOR FILING DATE: 1998-11-19
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1196
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-804-357-1
```

```
Query Match      53.1%; Score 906.4; DB 9; Length 1196;
Best Local Similarity 84.9%; Pred. No. 2.1e-261;
Matches 1015; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 244 AAATCAGTGTGGAGCTCGAAGAAATAAAGTTGTGGGTATCTCTGATGATGCAAAATGCT 303
DB 1 AGATCAGTGTGGGCTCGAAGAGTCAAGTTGCTGGGCATCTCTGATGATGCAAAATGCT 60
QY 304 TCATTTATTTTATTATGGAAGTCTCCAAAAGCAGTAGCCAAAGAAAAAATGGAAGGGG 363
DB 61 TCATTTATTTGATTGTGGAAGTCTCCAAAAGCAGTAGCCAAAGAAAAAATGGAAGGGAG 120
QY 364 AAGTAATTAATACCAAGAGAGAGTTCTGGAAGATATCTACCCCTCCCGAGGCATACCTGA 423
DB 121 GAGTAATTAATCCGAAAGAGAGAGTTCTGGAAGCACCACAGCACTCCCGGSCATACCTGA 180
QY 424 ACCGAGAGCAAGAGAGCTGAACCGGAGTAGTACAACCCCATCTCTCAGCATGCTGACCAACC 483
DB 181 ACAGGGAACAGAGAGAGTGAACAGGTGTAACATCCCATCTTGAACAGGGTGGCCCAATC 240
```

Qy	484	AGACGGGGAGGGGAGAGCTCTCCAAATATAGCCATCTGAACCTACTCGGACACCTGACC	543
Db	241	AGACAGGGGAGCTAGCCACATCTCCAAACACAAGTCACCTGAGCTATTGTGAACAGACT	300
Qy	544	TGAGGGTCACGTCGGTTCAGGGTTTANCAACTTGGCGGACAGATTAAAGACTTTC	603
Db	301	CGACGGTCATCAGACTGTCACAGATTTTAAATATCTGCCGACAGATTTAAGACTTTC	360
Qy	604	TGCTGTATTGTGAGATGCCGAATTAATTCATCTGCTTTATAGATCAACCGGATAAGTGTCAA	663
Db	361	TCCTGTATTGTGAGATGCCGAATTAATTCATCTGCTTTATAGATCAACCGGATAAGTGTCAA	420
Qy	664	AGAAACCTTCTCTGCTGGCAGTAATTAAGTCCTCACTCCACATTTTGGCCAGAGCAAG	723
Db	421	AGAAGCCCTCTTACTATTGGCGATAAAGTCCCTCATTTCCACATTTTGGCAGAGGCAAG	480
Qy	724	CAATCCGGGAATCTCTGGGGCAAGAAAGCAACGAGGAAACCAACAGCGTGGTGGAGTCT	783
Db	481	CAATTCGGGAGCTCTTGGGGCGGAAACCAACCGTAGGGAACCAAGACTAGTAGGGTCT	540
Qy	784	TCTGCTGGCCAGACACCCCAAGAGACAAACCCCGACCTTCAGATATGCTCAAT	843
Db	541	TCCTGTTGGCCAGACACCCCAAGAGACAAACCCCGACCTTACCTTGGCATGCTTAAGT	600
Qy	844	TTGAGACTGAGAAGCACCAAGACATCTTAATGTTGAACTACAGAGACACTTTCTTCAACT	903
Db	601	TTGAGAGTGACAAAGCACCAAGACATCTCATGTGAACTATAGAGACACATTTCTCAACC	660
Qy	904	TGCTCTGAAGGAAGTCTGTTTCTCAGGTGGGTAACTTCTGCTGCCAGACACTGAGT	963
Db	661	TGTCCTGAAAGGAAGTCTGTTTCTTAGTGGGTGAGCATCTCTGTCAGAGCCAGAGT	720
Qy	964	TTGTTTCAAGGGCGATGAGCATGTTTCTGTAACACCCATCATCTCTGAAATTAATGA	1023
Db	721	TTGCTTCAAGGGCGATGATGAGTGTTTGTGAACACCCATCATCTTAAATTAATGA	780
Qy	1024	ATAGTTATCCAGACCAAGCCAAAGATCTCTTCATAGTGTGATGCCAATGCTG	1083
Db	781	ATAGCTTATCCAGAGCAAGCCAAAGACTTGTTCATAGTGTGAGTATCCCAATGCTG	840
Qy	1084	GACCTCATCGGGATAAGAACTGAAGTACTACATCCCAAGATGTTTACTCTGGGCTCT	1143
Db	841	GGCTCACCGGATAAGAACTGAAGTACTACATCCCAAGATGTTTACTACCGGGCTCT	900
Qy	1144	ACCCACCTATGCAAGGGGAGGGGTTCTCTACTCCGGCCACTGGCCCTGAGGCTGT	1203
Db	901	ACCCACCGTATGCGGGGGTGGTGGATTCCTGTACTCCGGCCCTTGGCTTGGAGCTGT	960
Qy	1204	ACCATATCACTGACAGGTCCTCTCTACCCCATTTATGATGAGTTCATGGAATGTGCC	1263
Db	961	ACAGTGGGACTAGCGGGTCCATCTCTACCTTATGATGATGTTTATACGGGAATGTGCC	1020
Qy	1264	TTCAGAACTCGGCCTGTTCCAGAGAAACACAAAGGCTTCAGACATTTGATATCGAGG	1323
Db	1021	TTCAGAACTCGGCCTGTTCCAGAGAAACACAAAGGCTTCAGACATTTGATATGGAAG	1080
Qy	1324	AGAAAAACAAAAATAACATCTGCTCCCTATGTAGATCTGATGTTAGTACATAGTAAAAAC	1383
Db	1081	AGAAAAATAAGAAAAATATTGTCTCATATAGACCTAATGTTAGTACATAGCAGAAAAC	1140
Qy	1384	CTCAAGAGATGATTGATATTGGTCTCAGTTGCAGAGTGTCTCAATTTAAATGCTAA	1439
Db	1141	CTCAAGAGATGATTGATATTGGTCTCAGTTGCAAGTCTCAATTTAAATGCTGA	1196

RESULT 7

US-09-804-006-1
; Sequence 1, Application US/09804006
; Patent No. US20020119517A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.

```
QY 1024 ATAGTTTATCCAGACCAAGCCAAAGATCTCTTCATAGTGTATGATCCACAATGCTG 1063
DB 781 ATAGCTTTATCCAGAGCAAGCCAAAGACTTGTTCATAGTGACGTGATCCACAATGCTG 840
QY 1084 GACCTCATCGGGATAGAAGCTGAAGTACTACATCCAGAACTTGTTACTCTGGCTCT 1143
DB 841 GGCCTACCGGGATAGAAGCTGAAGTACTACATCCAGAACTCTTACACCGGCTCT 900
QY 1144 ACCACCTCATCGAGGGGAGGGGGTTCCTTACTCCGGCCACCTGGCCCTGAGGCTGT 1203
DB 901 ACCACCTCATCGAGGGGAGGGGGTTCCTTACTCCGGCCACCTGGCCCTGAGGCTGT 960
QY 1204 ACCATATCACTGACCAAGTCCATCTACGCCATCATGACGCTTATATCTGGAATGTC 1263
DB 961 ACAGTCGAGTACCGGGGTCCTCTACCTCTATCTACCTATTTATGATGATTTATACGGGAATGTC 1020
QY 1264 TCCAGAACTCGGCTCTGTTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATCGAGG 1323
DB 1021 TCCAGAACTCGGCTCTGTTCCAGAGAAACACAAAGGCTTCAGGACATTTGATATGAG 1080
QY 1324 AGAAACAAATAACATCTCTCTCTATGATCTGTATGATCTGTATGATAGTATAGAAAC 1383
DB 1081 AGAAACAAATAAGAAATATTTGTTCTATATAGACCTATGTTAGTATAGACAAAC 1140
QY 1384 CTCAGAGATGATGATGATTTGGTCTCAGTTGCAGAGTCTCATTTAAATGCTAA 1439
DB 1141 CTCAGAGATGATGATGATTTGGTCTCAGTTGCAGAGTCTCATTTAAATGCTGA 1196

RESULT 8
US-09-804-357-3
; Sequence 3, Application US/09804357
; Patent No. US20010024808A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/109001
; CURRENT APPLICATION NUMBER: US/09/804,357
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/195,896
; PRIOR FILING DATE: 1996-11-19
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1191
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-804-357-3

Query Match 53.0%; Score 904.6; DB 9; Length 1191;
Best Local Similarity 85.0%; Pred. No. 7.1e-261;
Matches 1012; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

QY 246 ATGAGTGTGGAGCTCGAAGATAAAGTTGTTGGGTATCTCTGATGATGCAAAATGCTTTC 305
DB 1 ATGAGTGTGGGGCTCGAAGATCAAGTTGCTGGCATCTCTGAGTATGCAAAATGCTTTC 50
QY 306 ATTTATTTTATTTATGAACTCTCCAAAGCAGTAGCCAGAAATAATGAAAGGGGAA 365
DB 61 ATTTATTTGATTTGGAAGTCTCCAAAGCAGTAGCCAGAAATAATGAAAGGGGAA 120
QY 366 GTAAATATACCCAAAGAGAGTTCTCGAAGATATCTACCCCTCCCGAGGACTACTGGAAC 425
DB 121 GTAAATATCCGAAAGAGAGTTCTCGAAGCCACCCAGCACTCCCGGCGCATCTGGAAC 180
QY 426 CGAGAGCAAGAGAGTGAACCGGCGAGTACAAACCCCATCTCTGAGCATCTGAGCAACCA 485
```

```
DB 181 AGGCAACAGAGAGAGCTGAACAGCTGTACAATCCCATCTTGAACAGGTGGCCAAATCAG 240
QY 486 ACGGGGAGGGGGGCGAGGCTCTCCAATATAGGCATCTGAACCTACTCGCAACCTGACCTG 545
DB 241 ACGGGGAGCTAGCCACATCTCCAAACACAAAGTCCCTGAGCTATTGTGAACAGAGCTG 300
QY 546 AGGTCAGCTGGTGGTTACGGGTTTAAACAACTTGGCGGACAGATTTAAAGACTTTTCG 605
DB 301 ACGTCTATGACAGCTGTGACAGATTTTAAATATCTGGCGGACAGATTTAAAGACTTTTC 360
QY 606 CTGTATTGACATCCGCAATTTATTCATCTGCTTATAGATCAGCCGGAATAAGTGTGCAAG 665
DB 361 TTGTATTGACATCCGGAATTTACTCGCTGCTATAGATCAACGGAAGAATAAGTGTGCA 420
QY 666 AAACCTTTCTTTGTTGCTGGGATTAAGTCCCTCACTCCATTTTGCAGAGCAAGCA 725
DB 421 AAGCCCTTCTTACTATTGGCGATTAAGTCCCTCACTCCATTTTGCAGAGCAAGCA 480
QY 726 ATCCGGGATCTCTGGGCGCAAGAAAGCAACGAGGGAACCAACGGTGGTCCGAGTCTTC 785
DB 481 ATCCGGGATCTCTGGGCGGAGAAACCAACGTTAGGGAACCAAGAGTAGTGAGGCTTTC 540
QY 786 CTGCTGGGCCAGACACACCCAGAGGAAACCAACCCGACCTTTTCAGATATGCTGAATTT 845
DB 541 CTGTTGGGCAAGACACCCCGAGAGGAAACCAACCCCTGACCTTTCCGACATGCTTAAGTT 600
QY 846 GAGAGTGAAGAGCAACCAAGACATTTCTATGTGGAACCTACAGAGACACTTTCTCAACTG 905
DB 601 GAGAGTGAAGAGCAACCAAGACATCTCATGTGGAACCTATAGAGACACTTTCTCAACTG 660
QY 906 TCTCTGAAGGAAGTCTGTTTCTCAGGTGGGTAAAGTACTTCTCCCGAGACACTGAGTTT 965
DB 661 TCCCTGAAGGAAGTCTGTTTCTTAGGTGGGTGAGCACTTCTCTCCAGAGCGAGATTT 720
QY 966 GTTTTCAAGGGCGATGACGATGTTTTTGTGAACCCCATCATCTCTGAAATTTACTTGAAT 1025
DB 721 GTCTTCAAGGGCGATGATGAGTGTGTTGTGAACCCCATCATCTCTTAAATTTACTTGA 780
QY 1026 AGTTTATCAAGACCAAGCCAAAGATCTCTTCAATAGGTGATGTGATCCCAATGCTGGA 1085
DB 781 AGCTTATCAAGAGCAAGCCAAAGACTTGTTCATAGGTGACGTGATCCCAATGCTGGG 840
QY 1086 CTTCTCGGGATAAGAGAGCTGAAGTACTATACATCCAGAAAGTTGTTTACTCTGGCTCTAC 1145
DB 841 CTTCAACCGGATAGAAACTGAAGTACTATACATCCAGAAAGTTTCTACACCGGCTCTAC 900
QY 1146 CACCTCTATGAGGGGAGGGGGTTCCTCTACTCCGGCCACCTGGCCCTGAGGCTGAC 1205
DB 901 CACCGTATGCGGGGGTGGTGGATTCCTGTACTCCGCCCCCTTGCCTTGAAGGCTGAC 960
QY 1206 CATATCACTGACAGGTCATCTCTACCCCATTCATGACGTTTATCTGGAATGTCCTT 1265
DB 961 AGTGGAGTACCGGGTCCATCTCTACCTATTGATGATGTTTATACGGGAATGTCCTT 1020
QY 1266 CAGAACTCGGCTCTGTTCCAGAGAAACAAAGGCTTCAGGACATTTGATATGAGGAG 1325
DB 1021 CAGAACTCGGCTCTGTTCCAGAGAAACAAAGGCTTCAGGACATTTGATATTTGAAGAG 1080
QY 1326 AAAAACAATAATACATCTGCTCTCTATGATGATCTGATGTTAGTACATAGTAAACCT 1385
DB 1081 AAAAATAAGAAAAATATTTGTTCTATATAGACCTAATGTTAGTACATAGCAAAAACT 1140
QY 1386 CAAGAGATGATTTGATTTTGGTCTCAGTTGAGAGTGTCTCATTTTAAATGTC 1436
DB 1141 CAAGAGATGATTTGATTTGTTCTGCTCAGTTGCAAGTCTCTAATTTTAAATGTC 1191

RESULT 9
US-09-804-006-3
; Sequence 3, Application US/09804006
; Patent No. US20020119517A1
; GENERAL INFORMATION:
```

APPLICANT: White, David
APPLICANT: Zhou, Jianghong
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: LEPTIN INDUCED GENES
FILE REFERENCE: 07334/126001
CURRENT APPLICATION NUMBER: US 09/804,006
CURRENT FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 09/292,228
PRIOR FILING DATE: 1999-04-15
PRIOR APPLICATION NUMBER: US 60/108,379
PRIOR FILING DATE: 1998-10-29
PRIOR APPLICATION NUMBER: US 09/150,857
PRIOR FILING DATE: 1998-09-10
NUMBER OF SEQ ID NOS: 17
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 1191
TYPE: DNA
ORGANISM: Mus musculus
US-09-804-006-3

Query Match
Best Local Similarity 53.08; Score 904.6; DB 10; Length 1191;
Matches 1012; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

246 ATGAGTGTGGACGTCGCAAGTAAGTCTGGTATCCTGATGATGCAAAATGCTTC 305
1 ATGAGTGTGGGCTCGAAGAGTCAAGTCTGCTGGGATCCTGATGATGCAAAATGCTTC 60
306 ATTATTTTATATGGAAGTCTCCAAAGCAGTAGCCAGCAAAATAATGGAAGGGAA 365
61 ATTATTTTATGTTGGAAGTCTCCAAAGCAGTAGCCAGCAAAATAATGGAAGGGAA 120
366 GATATATATCCCAAGAGAGTCTGCAAGATATCTACCCCTCCGAGGCGTACTGGAC 425
121 GATATATCCGAAGAGAGAGTCTGGAAGCCACCCAGCAGTCTCCCGGCGTACTGGAC 180
426 CGAGAGCAAGAGAGTCTGGAAGCCAGTCTGGAAGCCACCCAGCAGTCTGGAAGCCAG 485
181 AGGAGCAAGAGAGTCTGGAAGCCAGTCTGGAAGCCACCCAGCAGTCTGGAAGCCAG 240
486 ACGGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 545
241 ACGGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
546 AGGTACAGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 605
301 ACGGTACAGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
606 CTGTATTGAGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 665
361 TTGTATTGAGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
666 AAGCTTTCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 725
421 AAGCTTTCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 480
726 ATCCGGGAATCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 785
481 ATCCGGGAGTCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
786 CTGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 845
541 CTGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
846 GAGAGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 905
601 GAGAGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660
906 TCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 965
661 TCCCTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 720

966 GTTTTCAAGGGGATGACGATGTTTTCGTGAACCCCATCATCCTCGAATTAATTGAA 1025
721 GTCTTCAAGGGGATGACGATGTTTTCGTGAACCCCATCATCCTCGAATTAATTGAA 780
1026 AGTTTATCAAGACCAAGCCAAAGATCTCTTCATAGGTGATGTGATCCCAATGCTGGA 1085
781 AGCTTATCAAGACCAAGCCAAAGATCTCTTCATAGGTGATGTGATCCCAATGCTGGA 840
1086 CTTATCGGGATGAAGCTGAAGTACTACATCCCAAGATGTTGTTACTCTGGCTCTAC 1145
841 CTTACCGGGATGAAGCTGAAGTACTACATCCCAAGATGTTGTTACTCTGGCTCTAC 900
1146 CCACCTATGAGGGGAGGGGTTCTCTACTCCGCCACCTGGCCCTGAGGCTGAC 1205
901 CCACCTATGAGGGGAGGGGTTCTCTACTCCGCCACCTGGCCCTGAGGCTGAC 960
1206 CATATCACTGACCGGTCCTCTCTACCCCATGATGATGATGATGATGATGATGATGAT 1265
961 AGTGGGACTAGCGGGTCCATCTCTACCTATGATGATGATGATGATGATGATGATGAT 1020
1266 CAGAACTCGGCTCGTTCAGAGAACACAAAGCTTCAGACATTTGATATGAGGAG 1325
1021 CAGAACTCGGCTCGTTCAGAGAACACAAAGCTTCAGACATTTGATATGAGGAG 1080
1326 AAAAAACAAATAACATCTGCTCTATGATGATGATGATGATGATGATGATGATGATGAT 1385
1081 AAAAAACAAATAACATCTGCTCTATGATGATGATGATGATGATGATGATGATGATGAT 1140
1386 CAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1436
1141 CAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1191

RESULT 10
US-09-972-912-20
Sequence 20, Application US/09972912
Patent No. US20020110867A1
GENERAL INFORMATION:
APPLICANT: SOPPET, DANIEL R.
KUBEN, STEVEN M.
TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/972,912
FILING DATE: 10-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/049,022
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0620001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 857 base pairs
TYPE: nucleic acid
STRANDEDNESS: both

```

;
;   FILING DATE: <Unknown>
;   ATTORNEY/AGENT INFORMATION:
;   NAME: STEFFE, ERIC K.
;   REGISTRATION NUMBER: 36,688
;   REFERENCE/DOCKET NUMBER: 1488.0620001
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 371-2600
;   TELEFAX: (202) 371-2540
;   INFORMATION FOR SEQ ID NO: 26:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 480 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: both
;     TOPOLOGY: both
;   MOLECULE TYPE: cDNA
;   SEQUENCE DESCRIPTION: SEQ ID NO: 26:
US-09-972-912-26

Query Match      19.2%   Score 328.2;   DB 10;   Length 480;
Best Local Similarity 92.0%;   Pred. No. 8.9e-88;
Matches 402;   Conservative 0;   Mismatches 25;   Indels 10;   Gaps 5;

QY 103C TATCCAAAGACCAAGGAGTCTCTTCATAGGTGATGATCCACAAATGCTGGACCTC 1089
D 103C TATCCAAAGACCAAGGAGTCTCTTCATAGGTGATGATCCACAAATGCTGGACCTC 1089
D 20 TATTCAAAGACCAAGGAGTCTCTTCATAGGTGATGATCCACAAATGCTGGACCTC 79
QY 1090 ATCCGGATGAAGAGCTGAAGTACTACATCCAGAGAGTTGTTACTCTGCGCTCTACCCAC 1149
D 1090 ATCCGGATGAAGAGCTGAAGTACTACATCCAGAGAGTTGTTACTCTGCGCTCTACCCAC 1149
D 80 ATCCGGATGAAGAGCTGAAGTACTACATCCAGAGAGTTGTTACTCTGCGCTCTACCCAC 139
QY 1150 CCTATGCGAGGGGGGGGGTTCCTCTACTCCGGCCACCTGGCCCTGAGGCTGTACCATTA 1209
D 1150 CCTATGCGAGGGGGGGGGTTCCTCTACTCCGGCCACCTGGCCCTGAGGCTGTACCATTA 1209
D 140 CCTATGCGAGGGGGGGGGTTCCTCTACTCCGGCCACCTGGCCCTGAGGCTGTACCATTA 199
QY 1210 TCACCTGACGAGTCCATCTCTACCCCATGATGAGGTTTATCTGGAATGCGCTTCAGA 1269
D 1210 TCACCTGACGAGTCCATCTCTACCCCATGATGAGGTTTATCTGGAATGCGCTTCAGA 1269
D 200 TCACCTGACGAGTCCATCTCTACCCCATGATGAGGTTTATCTGGAATGCGCTTCAGA 259
QY 1270 AACTCGGCTCGTTCAGAGAAACACAAAGGCTTCAGGACATTTGATATCGAGGAGAAA 1329
D 1270 AACTCGGCTCGTTCAGAGAAACACAAAGGCTTCAGGACATTTGATATCGAGGAGAAA 1329
D 260 AACTCGGCTCGTTCAGAGAAACACAAAGGCTTCAGGACATTTGATATCGAGGAGAAA 319
QY 1330 ACAAAATAACATCTGCTCTATGATGATCTGATGATGATGATGATGATGATGATGATGAT 1389
D 1330 ACAAAATAACATCTGCTCTATGATGATCTGATGATGATGATGATGATGATGATGATGAT 1389
D 320 ACAAAATAACATCTGCTCTATGATGATCTGATGATGATGATGATGATGATGATGATGAT 379
QY 1389 GAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1446
D 1389 GAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1446
D 380 GAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 439
QY 1447 ACAAACTCAATTTKGSATNGAAGGGGTTTGTGATGATGATGATGATGATGATGATGATGAT 1506
D 1447 ACAAACTCAATTTKGSATNGAAGGGGTTTGTGATGATGATGATGATGATGATGATGATGAT 1506
D 440 ACAAACTCAATTTGGAATNGAAGGGGTTTGTGATGATGATGATGATGATGATGATGATGAT 499
QY 1507 TTAGAGTAACTTCTATTNANCA 1530
D 1507 TTAGAGTAACTTCTATTNANCA 1530
D 500 TTAGAGNGGTTCAAGTGGGTACA 523

RESULT 11
US-09-972-912-26
; Sequence 26, Application US/09972912
; Patent No. US20020110867A1
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R.
; RUBEN, STEVEN M.
; TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/972,912
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/049,022

```

```

;
;   FILING DATE: <Unknown>
;   ATTORNEY/AGENT INFORMATION:
;   NAME: STEFFE, ERIC K.
;   REGISTRATION NUMBER: 36,688
;   REFERENCE/DOCKET NUMBER: 1488.0620001
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 371-2600
;   TELEFAX: (202) 371-2540
;   INFORMATION FOR SEQ ID NO: 26:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 480 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: both
;     TOPOLOGY: both
;   MOLECULE TYPE: cDNA
;   SEQUENCE DESCRIPTION: SEQ ID NO: 26:
US-09-972-912-26

Query Match      19.2%   Score 328.2;   DB 10;   Length 480;
Best Local Similarity 92.0%;   Pred. No. 8.9e-88;
Matches 402;   Conservative 0;   Mismatches 25;   Indels 10;   Gaps 5;

QY 544 TGAGGTCACGTCGGTGGTTACGGGTTTAAACAATTTGCGGACAGATTTAAAGACTTTC 603
D 544 TGAGGTCACGTCGGTGGTTACGGGTTTAAACAATTTGCGGACAGATTTAAAGACTTTC 603
QY 604 TGCTGTAATTTGAGATGCCGAATTTATTCACCTGCTTATAGATCAGCCGGATAAGTGTC 663
D 604 TGCTGTAATTTGAGATGCCGAATTTATTCACCTGCTTATAGATCAGCCGGATAAGTGTC 663
D 62 TGCTGTAATTTGAGATGCCGAATTTATTCACCTGCTTATAGATCAGCCGGATAAGTGTC 120
QY 664 AGAAACCTCTTCTTGTGCTGGCGATTAAGTCCCTCACTCCACATTTTGCAGAGGCAA - 722
D 664 AGAAACCTCTTCTTGTGCTGGCGATTAAGTCCCTCACTCCACATTTTGCAGAGGCAA - 722
D 121 AGAAACCTCTTCTTGTGCTGGCGATTAAGTCCCTCACTCCACATTTTGCAGAGGCAA 180
QY 723 GCAATCGGGAATCTCGGGCCAGAGAAAGCAAGCAGGGAACCAACCGTGGTGGCAGTC 782
D 723 GCAATCGGGAATCTCGGGCCAGAGAAAGCAAGCAGGGAACCAACCGTGGTGGCAGTC 782
D 181 GCAATCGGGAATCTCGGGCCAGAGAAAGCAAGCAGGGAACCAACCGTGGTGGCAGTC 240
QY 783 TTCCTGCTGGCCAGACACCCAGAGAGCAACCCAGAGAGCAACCCAGACCTTTTCAGATATGCTGAAA 842
D 783 TTCCTGCTGGCCAGACACCCAGAGAGCAACCCAGAGAGCAACCCAGACCTTTTCAGATATGCTGAAA 842
D 241 TTCCTGCTGGCCAGACACCCAGAGAGCAACCCAGAGAGCAACCCAGACCTTTTCAGATATGCTGAAA 300
QY 843 -TTTGAGAGTGAGAGCAACCAAGACATTTCTATGTGGAA---CTACAGAGACACTTTCIT 898
D 843 -TTTGAGAGTGAGAGCAACCAAGACATTTCTATGTGGAA---CTACAGAGACACTTTCIT 898
D 301 TTTGAGAGTGAGAGCAACCAAGACATTTCTATGTGGAA---CTACAGAGACACTTTCIT 360
QY 899 CAACCTGTCTCT-CAAGGAAGTGTCTTCT---CAGGTGGGTAAAGTACTTCTCTGCCCA 953
D 899 CAACCTGTCTCT-CAAGGAAGTGTCTTCT---CAGGTGGGTAAAGTACTTCTCTGCCCA 953
D 361 CAANTTGTCTNTGGAAGAGAGTGTCTGTTTTTTCAGGTGGGTAAAGTACTTCTCTGCCCA 420
QY 954 GACACTGAGTTTGTGTTT 970
D 954 GACACTGAGTTTGTGTTT 970
D 421 GACATGAGTTTGTGTTT 437

RESULT 12
US-09-972-912-31
; Sequence 31, Application US/09972912
; Patent No. US20020110867A1
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R.
; RUBEN, STEVEN M.
; TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 10-Oct-2001
APPLICATION NUMBER: US/09/972.912
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/049,022
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0620001

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 31:

US-09-972-912-31
Query Match 16.7%; Score 284.8; DB 10; Length 303;
Best Local Similarity 99.3%; Pred. No. 7.2e-75;
Matches 286; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 585 GACAGATTAAAGACTTTCGCTGCTATTGAGATGCCGCAATTATTCACCTGCTTATAGAT 644
DB 1 GACAGATTAAAGACTTTCGCTGCTATTGAGATGCCGCAATTATTCACCTGCTTATAGAT 60
QY 645 CAGCCGGATAAGTGTCGAAAGAAACCTTCTGTTGCTGGCGATTAAAGTCCCTCACTCCA 704
DB 61 CAGCCGGATAAGTGTCGAAAGAAACCTTCTGTTGCTGGCGATTAAAGTCCCTCACTCCA 120
QY 705 CATTTCCGAGAGGCAAGCAATCCGGGAATCCTCGGGCCCAAGAAAGCAACGAGGGAAC 764
DB 121 CATTTCCGAGAGGCAAGCAATCCGGGAATCCTCGGGCCCAAGAAAGCAACGAGGGAAC 180
QY 765 CAAACGGTGTGGCAGTCTTCTGCTGGCCAGACACCCCGAGAGGACACCCCGGAC 824
DB 181 CAAACGGTGTGGCAGTCTTCTGCTGGCCAGACACCCCGAGAGGACACCCCGGAC 240
QY 825 CTTTCAGATATGCTGAAATTGAGAGTGAGAGCAACCAAGACATCTT 872
DB 241 CTTTCAGATATGCTGAAATTGAGAGTTAGAGCAACCAAGACATCTTCT 288
```

RESULT 13

US-09-972-912-11
Sequence 11, Application US/09972912
Patent No. US20020110867A1

GENERAL INFORMATION:

APPLICANT: SOPPET, DANIEL R.
RUBEN, STEVEN M.

TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:

ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 10-Oct-2001

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/972,912
FILING DATE: 10-Oct-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/049,022
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0620001

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Query Match 16.3%; Score 279; DB 10; Length 282;
Best Local Similarity 98.9%; Pred. No. 3.8e-73;
Matches 279; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 707 TTITGCCAGAGGCAAGCAATCCGGGAATCCTGGGGCCCAAGAAAGCAACGAGGGAACCA 766
DB 1 TTITGCCAGAGGCAAGCAATCCGGGAATCCTGGGGCCCAAGAAAGCAACGAGGGAACCA 60
QY 767 AACGGTGTGGCAGTCTTCTGCTGGCCAGACACCCCGAGAGGACACCCCGGACCT 826
DB 61 AACGGTGTGGCAGTCTTCTGCTGGCCAGACACCCCGAGAGGACACCCCGGACCT 120
QY 827 TTCAGATATGCTGAAATTTGAGAGTGAGAGCAACCAAGACATTTCTTATGGAACCTACAG 886
DB 121 TTCAGATATGCTGAAATTTGAGAGTGAGAGCAACCAAGACATTTCTTATGGAACCTACAG 180
QY 887 AGACACTTTCTCAACTTCTCTCTGAGGAAGTGTCTTCTCAGTGGGTGAAGTACTTC 946
DB 181 AGACACTTTCTCAACTTCTCTCTGAGGAAGTGTCTTCTNAGTGGGTGAAGTACTTC 240
QY 947 CTGCCCAGACACTGATTTGTTTTCAGGGCGATGACGATGT 988
DB 241 CTGCCCAGACACTGATTTGTTTTCAGGGCGATGACGATGT 282
```

RESULT 14

US-09-972-912-34

Sequence 34, Application US/09972912
Patent No. US20020110867A1

GENERAL INFORMATION:

APPLICANT: SOPPET, DANIEL R.
RUBEN, STEVEN M.

TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:

ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 10-Oct-2001

```
? CLASSIFICATION: <Unknown>
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 09/049,022
? FILING DATE: <Unknown>
? ATTORNEY/AGENT INFORMATION:
? NAME: STEFFE, ERIC K.
? REGISTRATION NUMBER: 36,688
? REFERENCE/DOCKET NUMBER: 1488.0620001
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 371-2600
? TELEFAX: (202) 371-2540
? INFORMATION FOR SEQ ID NO: 34:
? LENGTH: 282 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: both
? TOPOLOGY: both
? MOLECULE TYPE: cdna
? SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-09-972-912-34

Query Match      16.3%   Score 279;   DB 10;   Length 282;
Best Local Similarity 98.9%   Pred. No. 3.8e-73;
Matches 279;   Conservative 0;   Mismatches 3;   Indels 0;   Gaps 0;

QY 707 TTTTGGCAGAGGCAAGCAATCCGGGAATCCTGGGGCCAAAGAAAGCAAGCAGGGAACCA 766
Db 1 TTTTGGCAGAGGCAAGCAATCCGGGAATCCTGGGGCCAAAGAAAGCAAGCAGGGAACCA 60

QY 767 AACGGTGGTGGAGTCTTCTGCTGGCCAGACACCCCGAGAGGACACCCCGACCT 826
Db 61 AACGGTGGTGGAGTNTTCTGCTGGCCAGACACCCCGAGAGGACACCCCGACCT 120

QY 827 TTCAGATATGCTGAAATTTGAGAGTGAGAGACCAAGACATCTTAITGTGGAATACAG 886
Db 121 TTCAGATATGCTGAAATTTGAGAGTGAGAGACCAAGACATCTTAITGTGGAATACAG 180

QY 887 AGACACTTTCTCAACTTG-CTCTGAGGAAGTGCCTGTTCTCAGTGGGTAAGTACTTC 946
Db 181 AGACACTTTTCAACTTGCTCTGAGGAAGTGCCTGTTCTCAGTGGGTAAGTACTTC 240

QY 947 CTGCCAGACACTCAGTTCCTTTCAAGGGCGATGACGATGT 988
Db 241 CTGCCAGACACTGAGTTTG-TTTCAAGGGCGATGACGATGT 282
```

RESULT 15

```
US-09-972-912-36
? Sequence 36, Application US/0972912
? Patent No. US20020110867A1
? GENERAL INFORMATION:
? APPLICANT: SOPPET, DANIEL R.
? RUBEN, STEVEN M.
? TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE
? NUMBER OF SEQUENCES: 42
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
? STREET: 1100 NEW YORK AVENUE, SUITE 600
? CITY: WASHINGTON
? STATE: DC
? COUNTRY: US
? ZIP: 20005-3934
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/972,912
? FILING DATE: 10-Oct-2001
? CLASSIFICATION: <Unknown>
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 09/049,022
```

```
? FILING DATE: <Unknown>
? ATTORNEY/AGENT INFORMATION:
? NAME: STEFFE, ERIC K.
? REGISTRATION NUMBER: 36,688
? REFERENCE/DOCKET NUMBER: 1488.0620001
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 371-2600
? TELEFAX: (202) 371-2540
? INFORMATION FOR SEQ ID NO: 36:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 428 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: both
? TOPOLOGY: both
? MOLECULE TYPE: cdna
? SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-972-912-36

Query Match      15.7%   Score 267.4;   DB 10;   Length 428;
Best Local Similarity 99.6%   Pred. No. 1.6e-69;
Matches 268;   Conservative 0;   Mismatches 1;   Indels 0;   Gaps 0;

QY 942 ACTTCTGCCCCAGACACTGAGTTTCTTTCAAGGGCGATGACGATGTTTTCGTGAACACC 1001
Db 1 ACTTCTGCCCCAGACACTGAGTTTCTTTCAAGGGCGATGACGATGTTTTCGTGAACACC 60

QY 1002 CATCATCTCCTGAATTACTTGAATAGTTTATCCAGACCAAGCCAAAGATCTCTTCATA 1061
Db 61 CATCATCTCCTGAATTACTTGAATAGTTTATCCAGACCAAGCCAAAGATCTCTTCATA 120

QY 1062 GGTGATGATCCCAATGCTGGACCTCATCGGGATAGAAGCTGAAGTACTACATCCA 1121
Db 121 GGTGATGATCCCAATGCTGGACCTCATCGGGATAGAAGCTGAAGTACTACATCCA 180

QY 1122 GAAGTGTGTTTACTCTGGCCTCTACCCACCTATGCAGGGGGAGGGGTTCCTCTACTCC 1181
Db 181 GAAGTGTGTTTACTCTGGCCTCTACCCACCTATGCAGGGGGAGGGGTTCCTCTACTCC 240

QY 1182 GCCACCTGGCCCTGAGGCTGTACCATAT 1210
Db 241 GCCACCTGGCCTGAGGCTGTACCATAT 269
```

Search completed: October 20, 2003, 22:44:25
Job time : 473.998 secs